Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Vegetal

Phylogenetic insights of *Avena* genus based on genomic analysis of 45S rDNA molecular organization

Joana Rita António Rodrigues

Dissertação

Mestrado em Biologia Molecular e Genética

2014

Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Vegetal

Phylogenetic insights of *Avena* genus based on genomic analysis of 45S rDNA molecular organization

Dissertação orientada por

Doutora Manuela Silva

(Centro de Botânica Aplicada à Agricultura, Instituto Superior de Agronomia da Universidade de Lisboa)

Doutor Jorge Marques da Silva

(Departamento de Biologia Vegetal, Faculdade de Ciências da Universidade de Lisboa)

Joana Rita António Rodrigues

Dissertação

Mestrado em Biologia Molecular e Genética

2014

"The main thing in my kind of scientific work is to be able to fiddle with a thing, go on fiddling with it and fiddle, fiddle, until everyone else has given up."

Maurice Wilkins (Physicist, 1916-2004)

Agradecimentos

Irei começar por agradecer à professora Manuela Silva por me ter aceitado novamente na sua equipa. Por ter proposto uma linha de trabalho e perante os resultados obtidos, termos ido reformulando e/ou alterando significativamente os objectivos iniciais mais que uma vez. Por me deixar "testar só mais uma coisa" ou "ver o que são aquelas bandas" mesmo quando tal não era necessário. Por me ter dado total liberdade para "brincar" com o que tinha. Por todos os comentários, sugestões e horas passadas a discutir resultados ou simplesmente a ouvir-me divagar sobre o que tudo poderia significar. Mas principalmente por acreditar que eu seria capaz de concluir este trabalho mesmo quando eu não conseguia imaginar o fim do mesmo. O meu muito obrigado.

À professora Wanda Viegas, por me receber tão bem no seu laboratório. Pelas reuniões ao fim de tarde em que se decidiu "o melhor é deixar isto por agora e pegar noutra coisa" por ter mencionado naquelas escadas "então e o que se sabe do IGS? Vá estudar e daqui a uns dias diga coisas" quando a ideia de recomeçar naquele momento era assustadora. Pelos "bons-dias" mesmo quando eu estava lá no cantinho. E por me fazer sentir que tinha algo realmente interessante, quando eu só conseguia ver um alinhamento. Muito obrigado.

Ao professor Jorge Silva por me pedir para ler algo inicial e elogiar o modo como estava escrito, quando a ideia de escrever em inglês e redigir o primeiro artigo parecia inatingível. Muito obrigado.

Aos colegas do lab, por todas as dicas e conselhos, mas também pelas pausas para estufa/café/lanche… Por pararem o que estavam a fazer para olhar para as minhas coisas. O meu muito obrigado.

A todos os amigos que tiveram sempre uma palavra de apoio e um plano para desanuviar. Mas essencialmente à Sónia Pereira pelos fins de tarde a consumir açúcar pelos recantos de Lisboa, pelos mails intermináveis, por me dizer "tu consegues!" e "está quase" mas também "para quê que isso serve?" e "e se não der?", pelo optimismo e pessimismo na hora certa. À Adriana Resende por estar tão perto mesmo a 300kms de distância, pelas chamadas aos fins de tarde acompanhando-me no caminho para casa depois de mais um dia, por partilhar do "em modo de tese" nos bons e maus momentos. À Joana Marcelino pelos cafés depois de jantar, pelos dias que passámos enclausuradas na salinha na escrita, por partilhar o sentimento de estar *overwhelmed* com informação e por mudar o tema de conversa para "então e jogos/séries?". Muito obrigado por me manterem mentalmente sã.

À família por acreditar e apoiar incondicionalmente apesar dos "o quê que tu fazes mesmo?". Muitíssimo obrigado.

Resumo

O género *Avena* é constituído por 26 espécies com diferentes níveis de ploidia. Neste género existem diplóides com genoma AA ou CC e hibridação entre estas espécies originou os alopoliplóides com genoma AACC e posteriormente AACCDD. Existem ainda autotetraplóides com genoma AABB, uma vez que tanto o genoma B como o D derivaram do genoma A. Englobando espécies com diferentes níveis de ploidia e constituições genómicas diversas, este género é interessante para estudos filogenéticos, mas é igualmente importante economicamente, visto que várias espécies são cultivadas para alimentação.

O DNA ribossomal (rDNA), fundamental para a viabilidade celular, é extremamente conservado entre organismos tendo sido extensivamente utilizado para análise filogenética. Em *Avena*, tanto as sequências de rDNA 5S como os *internal transcribed spacers* (ITS) do rDNA 45S, foram já sequenciadas e utilizadas para tais estudos. No entanto, devido ao seu tamanho reduzido, estas regiões nem sempre permitem uma distinção clara das relações entre diferentes espécies. Nestes casos, o *5' external trascribed spacer* (5' ETS), parte do *intergenic spacer* (IGS) da subunidade 45S, tem vindo a originar resultados mais relevantes. Contudo, em *Avena* esta região ainda não foi estudada em detalhe. Por outro lado, restruturações de *loci Nor* (*nucleolar organizer regions*), as regiões dos cromossomas onde milhares de subunidades 45S de rDNA se encontram repetidas, são comuns em espécies poliplóides e têm sido igualmente descritas no género *Avena.*

Deste modo, este trabalho teve como objectivo avaliar a presença de polimorfismos no IGS das sequências ribossomais 45S entre espécies de aveia, particularmente através da caracterização da sua extremidade 5' ETS, aferir a utilidade desta região para análise filogenética e inferir a evolução do rDNA 45S no género *Avena*. Para isso foram analisadas as seguintes espécies: *A. strigosa* (AA), *A. ventricosa* (CC), *A. eriantha* (CC), *A. barbata* (AABB), *A. murphyi* (AACC), *A. sativa* (AACCDD) e *A. sterilis* (AACCDD). A análise filogenética foi ainda realizada através do estudo de sequências repetitivas dispersas no genoma, como retrotransposões e microssatélites, utilizando as metodologias *inter-retrotransposon amplified polymorphism* (IRAP) e *r*etrotransposon-*microsatellite amplified polymorphism* (REMAP).

A amplificação do IGS das referidas espécies com distintas constituições genómicas permitiu verificar que não existem variações entre indivíduos, tendo sido no entanto detectados polimorfismos interespecíficos nesta região, tanto no número como na dimensão dos fragmentos amplificados. Variações deste tipo já tinham sido observadas na espécie

hexaplóide *Avena sativa*, tendo sido proposto que se deviam a alterações no número de repetições de certas sequências presentes tanto no 5' ETS como no restante IGS.

Apesar de ser pouco comum, foi detectada variabilidade na ETS - região que é transcrita e bastante conservada – uma vez que foram observados marcados polimorfismos nestas regiões do genoma nas espécies de *Avena* analisadas. Surpreendentemente, verificou-se que as espécies diplóides com genoma C contêm 5' ETS de menores dimensões relativamente às restastes espécies e que um dos haplótipos estudados possui ETS de duas dimensões distintas, enquanto que em todas as restantes espécies foi apenas caracterizado um haplótipo. Adicionalmente, verificou-se que a sequência de menores dimensões amplificada em espécies diploides com o genoma C não está presente nas espécies poliplóides com este genoma (*A. murphyi*, *A. sativa* e *A. sterilis*). A sequenciação desta região revelou a presença de polimorfismos no domínio supostamente reconhecida pela RNA polimerase em *A. strigosa*, *A. sativa* e *A. sterilis*, assim como marcadas diferenças entre genomas na organização e número de motivos repetidos.

A análise filogenética das espécies estudadas foi realizada com base na caracterização das IGS do rDNA 45S e ainda utilizando marcadores envolvendo retrotransposões e microssatélites. O dendrograma construído com base nos padrões de bandas obtidos por IRAP e REMAP revelou uma distinção entre as espécies analisadas de acordo com o nível de ploidia e constituição genómica. A informação a obtida com base nas sequências do 5' ETS mostrou que, apesar da diferença de organização entre o genoma A e C, a distinção entre espécies é semelhante à obtida com os padrões de bandas de IRAP/REMAP. Assim, ficou provado que tanto regiões particularmente lábeis dos genomas, como retrotransposões e microssatélites, como regiões repetitivas codificantes, como as sequências de rDNA, poderão constituir excelentes ferramentas para aprofundar o conhecimento relativamente à biodiversidade no género *Avena*.

As sequências da extremidade 5' das ETS foram também comparadas com sequências publicadas para membros do mesmo sub-género, género e sub-família. Esta análise revelou que as sequências do genoma C apresentam uma organização mais similar à de outros níveis taxonómicos do que à organização observada nas sequências do genoma A de *Avena*. Este resultado vem apoiar a ideia de que o genoma C será o genoma ancestral de *Avena* do qual o genoma A divergiu. Assim, uma vez que a amplificação do 5' ETS das espécies diplóides com genoma C, revela um polimorfismo não observado nas espécies poliplóides com mesmo genoma, sugerimos que as alterações destas sequências que ocorreram ao longo da evolução do genoma A poderão constituir uma vantagem funcional. Esta hipótese é particularmente

interessante considerando as remodelações nos NORs usualmente observadas em espécies poliplóides, como silenciamento, perda e/ou homogeneização. Deste modo, os dados obtidos podem sugerir que em *Avena* a diferença entre este tipo de sequências pode estar correlacionada com a perda dos NORs do genoma C quando em condição poliploide, o que ainda não tinha sido proposto.

Como conclusão, os resultados obtidos neste estudo vieram clarificar a existência de polimorfismos no IGS dos genes ribossomais 45S de distintas espécies de *Avena*, particularmente nas extremidades 5' dos ETS dos dois genomas base de género, o que não era esperado dado que esta região é considerada bastante conservada. A análise filogenética destas regiões do genoma permitiu ainda sugerir o modo evolutivo particular de sequências de rDNA, contribuindo para a compreensão das relações filogenéticas das espécies deste género. Para além disso, a sugestão de relações entre a organização distinta das sequências de rDNA e a tendência para a sua eliminação diferencial ao longo da evolução, indica que o estudo do género *Avena* poderá constituir um modelo extremamente interessante para a caracterização dos mecanismos envolvidos nas reestruturações dos NORs que ocorrem em poliploidia.

Palavras-chave: *Avena*, DNA ribossomal 45S; 5' ETS; diversidade genómica; retrotransposões e microssatélites.

Abstract

The genus *Avena* comprises four distinct genomes organized in diploid, tetraploid and hexaploid species, constituting an interesting model for phylogenetic analysis. The aim of this work was to study genomic diversity of distinct *Avena* species – *A. strigosa, A. ventricosa, A. eriantha, A. barbata, A. murphyi, A. sativa* and *A. sterilis* – through the characterization of 45S rDNA IGS and using retrotransposon and microsatellite targeting markers.

The 5' External transcribed spacer (5'ETS) of the 45S ribosomal DNA has been promising in challenging phylogenetic studies but and has not been studied in detail in *Avena* genus. This work revealed the existence of length polymorphisms between the two dissimilar A and C genomes that can be ascribed to major differences in the number and organization of repeated motifs. Moreover, C-genome diploid species 5' ETS organization seems to retain more similarities with other Poaceae species than with *Avena* A-genome sequences, supporting the hypothesis of C-genome as being the ancestral *Avena* genome. However, the polyploid species with both genomes retain the A-genome 5'ETS organization and lack the C-genome sequences, suggesting their elimination during polyploid evolution as previously suggested.

The phylogenetic analysis obtained with 5' ETS sequences clustered the species studied according to ploidy and genomic constitution and the analysis of retrotransposons and microsatellites flanking sequences unravel similar results. Thus, our results support the use of the 5' ETS for phylogenetic inference in *Avena* species highlighting the evolutive pathways of ribosomal genes in plant species.

Key words: *Avena*; 45S ribosomal DNA; 5' ETS; genomic diversity; retrotransposons and microsatellites.

Table of contents

Appendix

List of tables and figures

Introduction

Oat is one of the most cultivated cereals worldwide and a valuable resource in several countries both for human consumption and animal feed (Strychar, 2011). Genus *Avena* L. belongs to the Poaceae family and comprises 26 species with different genomes and ploidy (Loskutov and Rines, 2011). Within this genus, the basic number of chromosomes is x=7 and there are diploid species with genome AA or CC; tetraploids with AABB or AACC-genomes and hexaploid species with genome AACCDD (Linares *et al.,* 1998; Katsiotis *et al.,* 2000). While the C-genome is quite different from the other ones (Nikoloudakis and Katsiotis, 2008), the B and D are almost identical to the A-genome (Katsiotis *et al.,* 1997; Linares *et al.,* 1998). Besides, since diploid species with B or D genomes have not been found, it seems that both diverged from an ancestral A' genome (Ananiev *et al.,* 2002). A model for hexaploid species evolution was proposed involving the hybridization of two diploid species, followed by chromosome duplication, originating an ancient allotetrapolyploid. Afterwards, this AACC genome underwent another hybridization event with an AA diploid species and subsequent chromosome duplication. The evolution of the tetraploid species with AABB genomes was an unrelated event that followed a diploid AA species autopolyploidization (for a review see Loskutov, 2008).

Considering the different genomes, levels of ploidy and that crosses between hexaploid species spawn fertile hybrids (Lawrence and Frey, 1975), this genus is an interesting model for phylogenetic studies. To clarify its evolution and diversity, several molecular markers have been covered such as: simple sequence repeats (SSR) or microsatellites (Li *et al.,* 2000; Fu *et al.,* 2007; Leišová *et al.,* 2007; He and Bjørnstad, 2012); restriction fragment length polymorphism (RFLP) (Goffreda *et al.,* 1992; Morikawa and Nishihara, 2009) and amplified fragment length polymorphism (AFLP) (Fu and Williams, 2008). Recently, inter-retrotransposon amplified polymorphism (IRAP) and retrotransposon-microsatellite amplified polymorphism (REMAP), that target mobile elements and SSR were also applied and proven useful to distinguish between *Avena* species (Tomás *et al.,* submitted). Those PCR-based techniques were initially implemented to identify barley cultivars (*Hordeum vulgare* L.) (Kalendar *et al.,* 1999) and have been used to assess genomic stability, to analyze genome constitution and evolution in hybrid species and to evaluate diversity, similarity and phylogenetic relationships within several genera (reviewed in Kalendar *et al.,* 2011; Bento *et al.,* 2013).

5S and 45S, the two ribosomal DNA subunits, are transcribed into ribosomal RNA (rRNA), the catalytic component of the ribosome crucial for cell viability. 5S rDNA

comprises tandems of conserved regions separated by non-transcribed spacers (Peng *et al.*, 2008), while the 45S rDNA is organized in tandem repeats of IGS-18S-ITS1-5.8S-ITS2-25S in the nucleolus organizer region (NOR) (Richard *et al.,* 2008). Conserved 18S and 25S coding sequences have not been studied in detail, however, the internal transcribed spacers (ITS1 and ITS2) as well as the ITS1-5.8S-ITS2 region have been sequenced for *Avena* phylogenetic analysis (Nikoloudakis *et al.,* 2008; Nikoloudakis and Katsiotis, 2008; Peng *et al.,* 2010; Rodionov *et al.,* 2005 and Tomás *et al.,* submitted). The ITS is easily amplified and evolves rapidly which makes it valuable for phylogenetic inference. However, to differentiate between closely related species and clarify phylogenetic relationships the study of a larger and more variable region – the external transcribed spacer (ETS), located in the intergenic spacer (IGS) – has also been useful (Linder *et al.,* 2000).

In 1994, Polanco and Pérez de la Vega sequenced the longest intergenic spacer of *Avena sativa* with 4098bp (Accession Number: X74820.1) and five different types of repeats were identified– A, B, C, D and E – each one present in two to 13 repetitions. Furthermore, the sequence TATAGTAGGG was reported as being *Avena'*s transcription initiation site (TIS), marking the beginning of the 5' external transcribed spacer where RNA polymerase I initiates transcription of ribosomal RNA and several pre-rRNA processing signals are also present (Sáez-Vasquez and Echeverría, 2008; Nazar, 2004). Inter and intraspecific variations in the IGS have been described in *Avena* by RFLP (Cluster and Allard, 1995; Jellen *et al.*, 1994 and Nikoloudakis *et al.,* 2008). However, there are opposing reports regarding the PCR amplification of this region. Polanco and Pérez de la Vega (1997) described intraspecific variations in IGS length of 58 varieties and 14 landraces of *Avena sativa* unraveled by the presence of two to six polymorphic IGS fragments out of 12 variants in each accession. Although none of the IGS polymorphisms were sequenced, the length variety was suggested to be due to differences in the copy number of repeats. Conversely, Nikoloudakis and colleagues (Nikoloudakis *et al.,* 2008) did not detect species differentiation through IGS amplification.

Molecular and cytogenetic methods to study ribosomal DNA (rDNA) have been applied extensively to reveal relationships between and within *Avena* species (Irigoyen *et al.,* 2002; Linares *et al.,* 1992, 1996 and 1998; Nikoloudakis *et al.,* 2008; Peng *et al.,* 2008 and 2010). The physical mapping of rDNA has been broadly studied, either by fluorescence *in situ* hybridization (FISH) or, to a lesser extent, by Ag-NOR. This silver-staining method allows the recognition of active NORs (Caperta *et al.*, 2002), while with FISH all NORs are detected. The data available by FISH in *Avena,* summarized in Table 1, revealed discrepancies in the

number of NORs reported, as for instance in the hexaploid species *Avena sterilis* where six, eight and 14 NOR loci have been described. Nonetheless, it is clear that throughout *Avena* evolution allopolyploidization was followed by a preferential decrease in the number and/or size of C-genome NORs, as mentioned by Nikoloudakis and Katsiotis (2008). In allopolyploids, rDNA restructuring events such as sequence homogenization, NOR silencing and NOR loss have been reported (Nieto Feliner and Rosselló, 2012; Pikaard, 2000; Maluszynska and Heslop-Harrison, 1993). However, the reason why NORs from one progenitor are preferentially eliminated or silenced is not entirely known. Still, translocations, unequal crossing-over and transposition could be implied (Maluszynska and Heslop-Harrison, 1993; Winterfeld et al., 2009). Considering the lack of information regarding this subject, particularly in *Avena*, the present work aimed to i) evaluate IGS length polymorphisms in different species; ii) characterize the 5' ETS diversity; iii) assess its usefulness for phylogenetic purposes in comparison with the study of retrotransposons and microsatellites as well as iv) enlighten *Avena* genus evolution pathways.

Table 1: Number and chromosome location of nucleolar organizer regions reported for selected *Avena* species.

Species	**Genome**	**NORs**	**Observation**	**Reference**
A. strigosa	AsAs	4	Similar size	Linares *et al.*, 1996
A. ventricosa	CvCv	6	2 majors; 4 minors	Badaeva *et al.*, 2010a
A. eriantha	CpCp	6	4 majors; 2 minors	Linares *et al.*, 1996
A. barbata	AABB	8	A-Genome: 4 majors; B-Genome: 2 majors; 2 minors	Irigoyen *et al.*, 2001
A. murphyi	AACC	4	A-Genome: Similar size	Linares *et al.*, 1996
		6	A-Genome: 4 majors; C-Genome: 2 minors	Fominaya *et al.*, 1995
A. sativa	AACCDD	6	A/D-Genome: Similar size	Linares *et al.*, 1998
		12	6 majors; 6 minors	Nikoloudakis and Katsiotis, 2008
		14	A/D-Genome: 6 majors; 2 minors C-Genome: 6 minors	Badaeva *et al.*, 2011
A. sterilis	AACCDD	6	Similar size	Tomás *et al.*, submitted
		8	A/D-Genome: 6 majors; C-Genome: 2 minors	Fominaya *et al.*, 1995
		14	A/D-Genome: 6 majors; 2 minors C-Genome: 6 minors	Badaeva *et al.*, 2011

Materials and Methods

The plant material used in this work was an old Portuguese line of *A. strigosa* (str, genome AA) from the Portuguese Vegetal Germplasm Bank (EAN 82, PRT005) and the remaining species came from the United States Department of Agriculture – National Small Grains Germplasm Research Facility: *A. ventricosa* (ven, genome CvCv – PI657337); *A. eriantha* (eri, genome CpCp – Clav9050); *A. barbata* (bar, genome AABB – PI367338); *A. murphyi* (mur, genome AACC – PI657606); *A. sativa* (sat, genome AACCDD – Clav8250) and *A. sterilis* (ste, genome AACCDD – PI267989).

At least three seeds were germinated and maintained in growth chambers at 16h/day (22ºC) and 8h/night (15ºC). Leaves from different one month old plants were frozen at -80ºC for further utilization and plants were then maintained in greenhouse until life cycle completion.

DNA extraction

DNA extraction was performed following the Citogene® DNA Cell & Tissue Kit (cat. No. CT-0150) from Citomed: approximately 30 mg of frozen tissue was grinded with mortar and pestle in liquid nitrogen and 300 µl cell lysis were added. The mixture was vortexed and incubated for 60 minutes at 65ºC. The tube was inverted few times after 30 and 60 minutes. After 1.5 µl of RNase A was added, the sample was inverted several times and incubated for 20 minutes at 37ºC. Finally the sample was cooled to room temperature, 100 µl of protein precipitation was added and the tube was vortexed for 20 seconds. Centrifugation at 16000 x g was carried for 3 minutes and the supernatant was transferred to a clean 1.5 ml tube containing 300 µl of cold 100% isopropanol. The tube was inverted several times and remained at room temperature for at least 30 minutes. Then the sample was centrifuged at 16000 x g for 1 minute and the supernatant was discarded. After drained on absorbent paper, 300 µl of cold 70% ethanol was added, the tube was inverted several times and centrifuged for 1 minute at 16000 x g. The supernatant was poured carefully and the sample was left to dry for 10 to 20 minutes. Afterwards, 50 µl of DNA hydration was added and left overnight at room temperature. DNA was quantified and stored at -20ºC.

PCR experiments

Each PCR was performed in 0.2 ml tubes and for a total volume of 20 µl, it was used: 1x PCR buffer[1]; 1.5 mM $MgCl_2$[1]; 0.1 mM dNTP's[2]; 0.25 to 1 µM of each primer; 1 U NZYTaq DNA polymerase[1]; 30 to 50 ng DNA template and Milli-Q H_2O for the remaining volume. PCR experiments were performed in at least three replicates for each primer combination. The IGS was amplified using the modified touchdown PCR (Don *et al.*, 1991): 94ºC for 4 minutes; 10 cycles of 94ºC for 1 minute; 65ºC for 30 seconds, decreasing 0.5ºC each cycle; 72ºC for 4 minutes; 25 cycles of 94º for 1 minute; 60ºC for 30 seconds; 72ºC for 4 minutes and 10 minutes of final extension at 72ºC. Amplification of partial ETS sequences was carried as followed: 94ºC for 4 minutes; 30 cycles of 94ºC for 1 minute; 57ºC for 45 seconds; 72ºC for 90 seconds; with 15 minutes of final extension at 72ºC. The complete ETS sequences were amplified with the same program but with 2 minutes in each extension step. The amplification of the ITS2 region was carried with the partial ETS program but with 60ºC in the annealing step. IRAP and REMAP were performed according to Kalendar *et al.*, (1999) using the following program: 94ºC for 2 minutes; 30 cycles of 94ºC for 30 seconds; 50ºC for 30 seconds increasing 0.5ºC each cycle; 72ºC for 2 minutes; with 10 minutes of final extension at 72ºC.

Table 2: Primers used for PCR reactions.

Primer designation	Sequence
Nikita*	5' – CGCTCCAGCGGTACTGCC – 3'
(CA)9G	5'– CACACACACACACACACAG – 3'
25S**	5' – GACGACTTAAATACGCGACGG – 3'
18S***	5' – AGACAAGCATATGACTACTGG – 3'
ETS1_fow	5' – TGTACCCCTCCTTCACAAGC – 3'
ETS1_rev	5' – CGAGGCTTCCTTGATAGCAC – 3'
ETS2_fow	5' – AAAACCCGTGCAGGAACTC – 3'
ETS2_rev	5' – CAAGCACTTGAAAGGCAACA – 3'
ETS3_fow	5' – GGACACTCAGCACGCCTTC – 3'
ETS3_rev	5' – ACACGGGTCCAAAGCTACTC – 3'
ETS4_fow	5' – TCGGTGTTTACATGTTCGAG – 3'
ITS3****	5' – GCATCGATGAAGAACGCAGC – 3'
ITS4****	5' – TCCTCCGCTTATTGATATGC – 3'

[1] Cat. No. MB00103
[2] Cat. No. MB08701

Note: Primers ETS1 to ETS3 were designed in primer3[3] based on the *A. sativa* sequence (Accession Number: X74820.1) and ETS4_fow based on the partial sequences of *A. ventricosa* (Accession Number: KM586761). * Baumel *et al.*, 2002. ** Polanco and Pérez de la Vega, 1994. *** designed based on the *Triticum aestivum* 18S sequence (Accession Number: X07841). **** White *et al*., 1990.

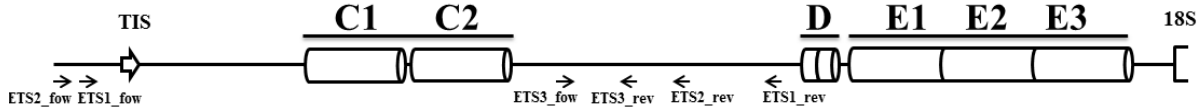


Figure 1. Representation of the primers designed for the 5' ETS amplification based on the *A. sativa* sequence (Accession Number: X74820.1). The sequence of the primer ETS4_fow does not exist in the *A. sativa* and is not represented. In *A. ventricosa* is located downstream the C repeat.

PCR products were separated in 1 or 1.7% agarose gels with 1 Kb Plus DNA Ladder (Invitrogene – cat. No. 10597-011) as molecular marker. Gels were detected with ethidium bromide and photographed using BIO-Rad GEL DOC 2000.

Sequence cloning

The amplified bands were isolated from the agarose gel, placed in a 1.5 ml tube and purified with NZYGelpure (cat. No. MB01102), following the manufacturer instructions: 3 times the gel weight in volume was added of binding buffer and the mixture was incubated at 60ºC for 10 minutes. The mixture was transferred to a spin column placed into a collection tube and centrifuged at 16000 x g for 1 minute. The flow-though was discarded and 500 µl of wash buffer added. A similar centrifugation was carried and the step repeated with 600 µl of the same buffer. The spin column was centrifuged again and placed into a clean 1.5 ml tube. 30 µl of elution buffer warmed at 37ºC were added to the center of the column and incubated at room temperature for 1 minute. One last centrifugation at 16000 x g was carried for 1 minute.

The purified fragments were quantified and ligated into the pCR®2.1 vector from Invitrogen (Cat. No. K2040-40). In a 0.2 ml tube, the ligation reaction was as follows: X[4] µl of fresh PCR product (less than 24 hours after amplification); 1 µl of 10X ligation buffer; 2 µl of pCR®2.1 vector (25 ng/µl), sterile water to a total volume of 9 µl and 1 µl T4 DNA ligase (4.0 Weiss units). The ligation reaction was incubated overnight at 14ºC and stored at-20ºC.

The ligation was cloned into NZY5α Competent cells (Cat. No. MB00401) as follows: the competent cells were thawed in ice and 67 µl were transferred into a 1.5 ml tube. 8 µl of

[3] Koressaar and Remm, 2007; Untergrasser *et al*., 2012

[4] $X\ ng\ PCR\ product = \frac{(Y\ bp\ PCR\ product PCR) \times (50\ ng\ pCR® 2.1\ vector)}{(size\ in\ bp\ of\ the\ \ pCR® 2.1\ vector \sim 3900)}$

the ligation was added and the tubes were gently mixed. The cells were incubated on ice for 30 minutes than placed for 40 seconds into a 42ºC water bath and immediately after transferred to ice for 2 minutes. 540 µl of SOC medium[5] at room temperature was added. The cells were incubated for 90 minutes at 37ºC and 250 rpm agitation. 50 and 100 µl of transformed cells were plated on LB agar plates[6] containing ampicillin (100 µg/ml) and X-Gal[7] (20 µg/ml) and incubated at 37ºC overnight.

For colony screening, PCR reactions were performed with primers M13 (5'-TGTAAAACGAGCCAGT-3' and 5'-CAGGAAACAGATGACC-3') with the following program: 94ºC for 3 minutes; 30 cycles of 94ºC for 45 seconds; 54ºC for 40 seconds; 72ºC for 90 seconds or 2 minutes; with 5 minutes of final extension at 72ºC. Selected colonies were grown overnight at 37ºC, 250 rpm in liquid LB medium[8] with ampicillin.

Plasmid purification was carried with NZYMiniprep (cat. No. MB01002) according to the manufacturer's instructions with minor modifications: 5 ml of the LB cultures were centrifuged for 2 minutes at 8000 x g and the supernatant was discarded. The cells were re-suspended in 250 µl of buffer A1 by vortexing the tubes. 250 µl of buffer A2 was added and the tubes slowly inverted several times, followed by 4 minutes of incubation at room temperature. 300 µl were added to the mixture and the tubes were inverted few times once more. Centrifugation was carried for 10 minutes at 11000 x g and the supernatant was transferred into a spin column in a collection tube. The samples were centrifuged at 11000 x g for 1 minute and the flow-through discarded. 500 µl of AY buffer, warmed to 50ºC, were added and the column was centrifuged for 1 minute at 11000 x g. The previous step was repeated with 600 µl of buffer A4. The spin column was centrifuged for 2 minutes at 11000 x g. The dried spin column was placed into a clean 1.5 ml tube and 50 µl of elution buffer at 50ºC were added. After 1 minute of incubation at room temperature, the tube was centrifuged for 1 minute at 11000 x g. The purified plasmid was stored at -20ºC.

The clones obtained were sequenced through the Sanger method[9] and the accession numbers of all sequences obtained are available in appendix I.

[5] 2% Tryptone; 0,5% Yeast Extract; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; 20 mM Glucose

[6] 1.0% Tryptone; 0.5% Yeast Extract; 1.0% NaCl; 15g/L Agar; pH 7.0 autoclaved. After cooled to ~50^{o}C, ampicillin was added and the medium plated.

[7] 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

[8] 1.0% Tryptone; 0.5% Yeast Extract. pH 7.0, autoclaved

[9] Stabvida (Campus FCT UNL Edifício Departamental de Química 2825-516 Caparica)

Sequence and phylogenetic analysis

The sequences were aligned using the ClustalW application (Thompson *et al.*, 1994; Larkin *et al.*, 2007) and adjusted manually using BioEdit Sequence Alignment Editor version 7.2.5 (Hall, 1999). Dotplots to identify repeats were constructed in Unipro UGENE (Okonechnikov *et al.*, 2012).

The Neighbour Joining (NJ) tree with 1000 bootstrap repetitions was obtained in MEGA version 6 (Tamura *et al.*, 2013). Since some sequences presented insertions, one standard sequence - the one with the highest similarity with the published sequence for *A. sativa* (Accession Number: X74820.1) - of each species was used for this analysis, instead of a consensus sequence. The banding profiles obtained from the IRAP and REMAP were transformed into binary matrices, where 1 and 0 correspond to the presence or absence of band, respectively. The Dice coefficient (Dice, 1945) was used to evaluate the matrices similarity in NTSYSpc version 2.02 (Rohlf, 2008). The Unweighted pair group method with arithmetic mean (UPGMA) in the same software was used to obtain the dendrogram. The bootstrap values for 1000 repetitions were obtained with WinBoot (Yap and Nelson, 1996). The images presented were processed with Photoshop CS2 version 9.02.

Results

45S rDNA genomic characterization

Intergenic spacer polymorphisms evaluation

The entire 45S rDNA intergenic spacer was assessed in all species in study through PCR amplification using two or three individuals of each species. No polymorphisms within species were observed (Figure 2A) and therefore only one individual per species was further analyzed. However, considerable IGS length polymorphism in the complete IGS between species can be easily observed (Figure 2B) in the 4000bp gel region. *Avena strigosa* presented the longest spacer with ~4000bp, while all the others have shorter variants. *A. ventricosa, A. barbata* and *A. sterilis* seem to have IGSs of similar dimension (~3600bp), as well as *A. eriantha, A. murphyi* and *A. sativa* (~ 3800bp). Moreover, an additional band with lower intensity can be seen in *A. eriantha* and *A. barbata.*

Furthermore, bands with shorter dimension than expected were detected with: (i) ~500bp in *A. strigosa*, *A. barbata*, *A. murphyi*, *A. sativa* and *A. sterilis;* (ii) ~800bp in *Avena sterilis* and (iii) ~1500bp in *A. murphyi*. The band presenting the highest intensity in *Avena sterilis* (~800bp) was sequenced and the consensus sequence obtained from two clones can be seen in the appendix II. The comparison with the published IGS sequence of *A. sativa* (Accession Number: X74820.1) revealed that this sequence comprises similarity with the end of the 25S subunit and the region from 3390bp to 4098bp corresponding to a partial region of the external transcribed spacer immediately upstream 18S and the beginning of the 18S subunit.

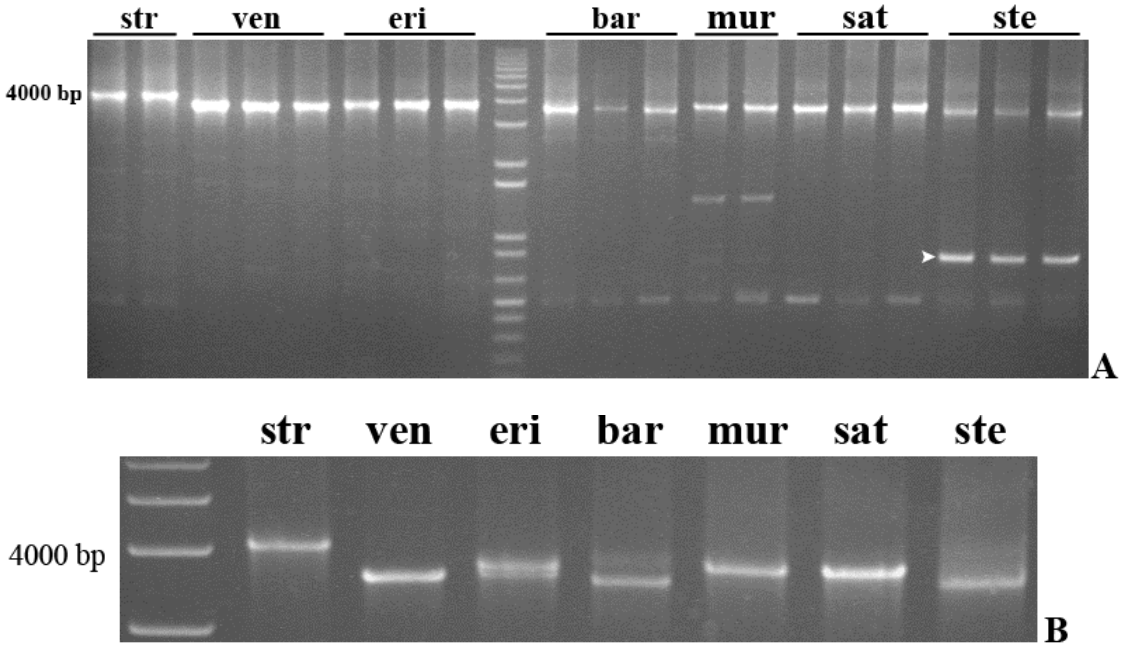


Figure 2. PCR amplification of the complete intergenic spacer (IGS) of two or three individuals per species (A) and one individual focused on the 4000bp region (B). (A) Lane 1 and 2 – *A. strigosa; 3 to 5 – A. ventricosa*; 6 to 8 – *A. eriantha*; 9 – molecular marker 1Kb+; 10 to 12 – *A. barbata*; 13 and 14 – *A. murphyi*; 15 to 17 – *A. sativa* and 18 to 20 – *A. sterilis*. (B) Lane 1 – molecular marker 1Kb+; 2 – *A. strigosa*; 3 – *A. ventricosa*; 4 – *A. eriantha*; 5 – *A. barbata*; 6 – *A. murphyi*; 7 – *A. sativa* and 8 – *A. sterilis*. The arrowhead marks the band sequenced.

5' ETS partial sequences analysis

The *A. sativa* 45S rDNA 5' external transcribed spacer (ETS) published is approximately 1750bp long, thus its amplification and direct sequencing to evaluate variation in the number of repetitive elements within this region is not easily feasible. Hence the first focus was to study the region from upstream of the TIS to the beginning of D repeat. Two forward primers were designed for this region: ETS1_fow located closer to the TIS and containing the motif CCCTCC identified previously (Polanco and Pérez de la Vega, 1994) that revealed high homology to several species of the Poaceae family; and ETS2_fow designed for a region 15bp upstream of the ETS1_fow expected to be useful to amplify the full 5' ETS including a larger portion of the region potentially recognized by the RNA polymerase I. Internal primers had to be designed as well, thus ETS1_rev and ETS2_rev were designed for a conserved region described previously (Polanco and Pérez de la Vega, 1994).

The expected fragments sizes amplified with the different primers combinations were as follows: ETS1_fow and ETS1_rev, 1089 bp; ETS2_fow and ETS2_rev, 977 bp; ETS1_fow and ETS2_rev, 942 bp and ETS2_fow and ETS1_rev, 1129 bp. Amplification with primer ETS1_fow of species with the A-genome - *A. strigosa*, *A. barbata* and *A. sterilis* - produced only the fragment with the expected dimension (~1100 bp and ~940 bp with ETS1_rev or ETS2_rev primer, respectively). On the other hand, *A. ventricosa* yielded a ~150 bp larger band and several smaller ones with much lower intensity when combined with primer ETS2_rev (Figure 3, lane 3) and a single ~150 bp larger band with primer ETS1_rev (Figure 3, lane 11). Amplification with primer ETS2_fow was expected to reveal only one fragment with the dimensions mentioned above. However, that band was only observed in *A. strigosa* (lane 6 and 14), *A. sterilis* (lane 9 and 17), *A. murphy* and *A. sativa* (not shown). *A. barbata* presented the expected fragment and other bands, all with much lower intensity than the other A-genome species (lane 8 and 16). *A. eriantha* amplification (not shown) revealed to be similar to *A. ventricosa*: a complete lack of amplification with primers ETS2_fow and ETS1_rev (lane 15) and an extremely faint fragment with ~3000 bp with primer ETS2_fow

and ETS2_rev (lane 7). So it seems that the ETS2_fow sequence is absent in the C-genome diploid species and uncommon in the genomes of *A. barbata*, although it has the A-genome.

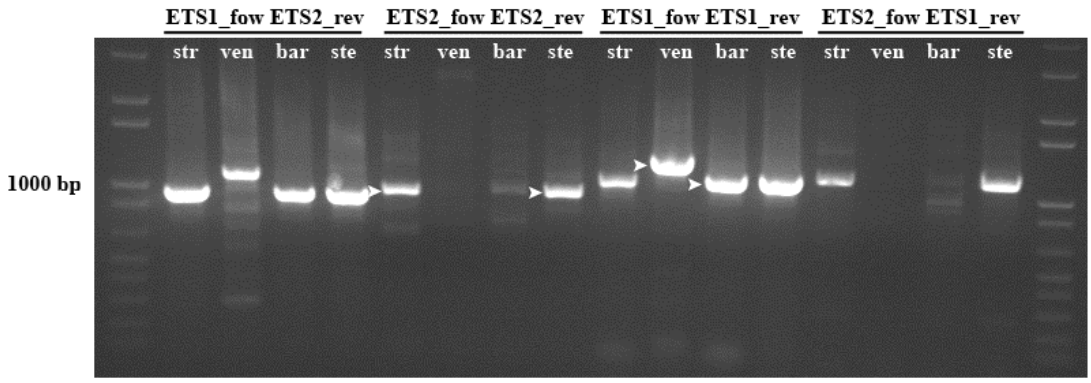


Figure 3. Parcial amplification of the 5' ETS. Lane 1 and 18 – molecular marker 1Kb+; 2 to 5 – primers ETS1_fow and ETS2_rev; 6 to 9 – primers ETS2_fow and ETS2_rev; 10 to 13 – primers ETS1_fow and ETS1_rev; 14 to 17 – primers ETS2_fow and ETS1_rev. *A. strigosa* (2,6,10,14); *A. ventricosa* (3,7,11,15); *A. barbata* (4,8,12,16) and *A. sterilis* (5,9,13,17). Arrowheads mark bands sequenced.

PCR amplification products obtained with primers to the ETS2 sequence in *A. strigosa*, *A. sativa* and *A. sterilis* and with primers to the ETS1 sequence in *A. ventricosa* and *A. barbata* were sequenced (arrowheads in Figure 3). The alignment of ETS1 and ETS2 sequences is available as supplemental material as well (Appendix III), including one clone from *A. ventricosa* (two clones were sequenced with 100% homology, thus only one is represented) and three from every other species. The overall similarity of all sequences with the referred published one from *A. sativa* was 66% in *A. ventricosa* and 94 to 99% in the remaining species. The blastn search aligned a 507-881bp region of *A. ventricosa* sequence with *A. sativa* published sequence domain that includes the E repeats, although the presence of E repeats in this region was not expected. Also, only one C repeat was detected in the *A. ventricosa* sequences. Regarding *A. sativa*, one clone sequence was shorter than the others obtained and its alignment revealed the lack of the C2 repeat as well.

Moreover, in most sequences obtained with ETS2 primers in *A. strigosa*, *A. sativa* and *A. sterilis*, insertions of a sequence with 30bp, or fragments of it, were detected surrounding the TIS. Figure 4 shows the alignment of all clones that showed variation in this region. A search for regulatory elements in these sequences performed in PLACE database (Higo *et al.*, 1999) revealed that 50% of motifs detected were related to promoter sequences. Single nucleotide polymorphisms (SNPs) were also detected up and downstream of the TIS, mainly in *A. ventricosa*. Additionally, a 10bp insertion starting in position +25 was also detected in the same species.

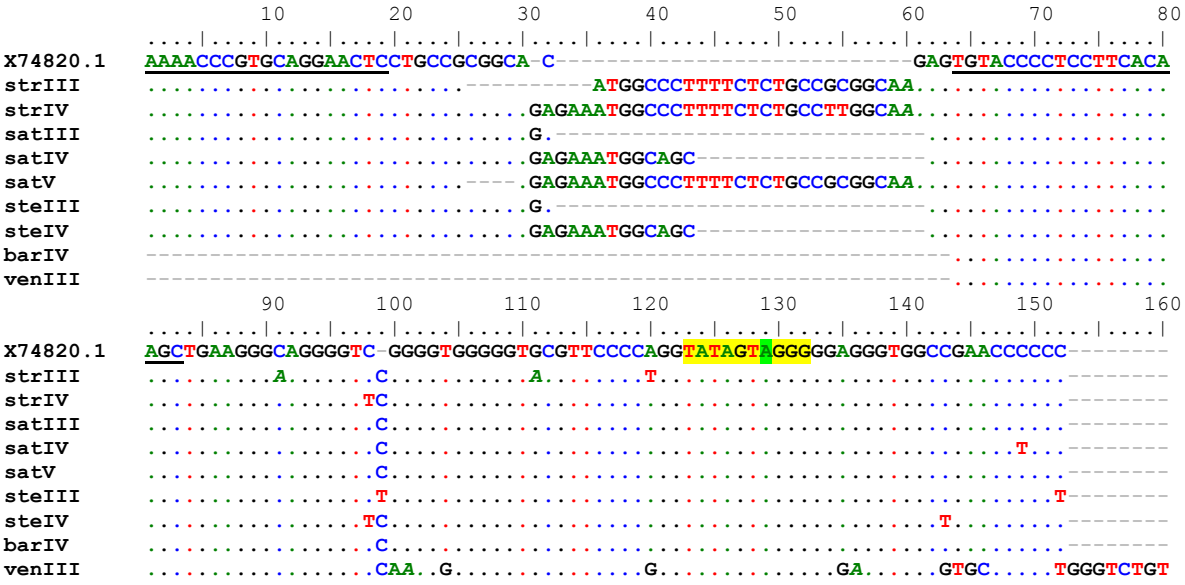


Figure 4. IGS sequences alignment surrounding the transcription initiation site (TIS). Primer sequence of ETS1_fow (from position 64 to 83) and ETS2_fow (from position 1 to 19) are underlined. TIS annotated in the published sequence (Accession Number: X74820.1) is highlighted in yellow. TIS position +1 is highlighted in green.

5' ETS complete sequences analysis

The data obtained allowed the characterization of the complete 5' ETS in all species studied. Primers ETS1_fow and 18S were used to amplify that region and all species with A-genome – *A. strigosa*, *A. barbata*, *A. murphyi*, *A. sativa* and *A. sterilis* – revealed 5' ETS with similar length since a single band with 1750bp was observed. C-genome diploid species – *A. ventricosa* and *A. eriantha* – showed shorter ETS sequences (~1500bp) and *A. eriantha* revealed an additional band in the 1400bp region (Figure 5). The presence of shorter ETS sequences in the C-genome diploid species is particularly interesting considering that *A. ventricosa* partial sequences are larger than expected. On the other hand, none of the polyploid species with C-genomes - *A. murphyi*, *A. sativa* and *A. sterilis* - has the shorter variant present in the C-genome diploids.

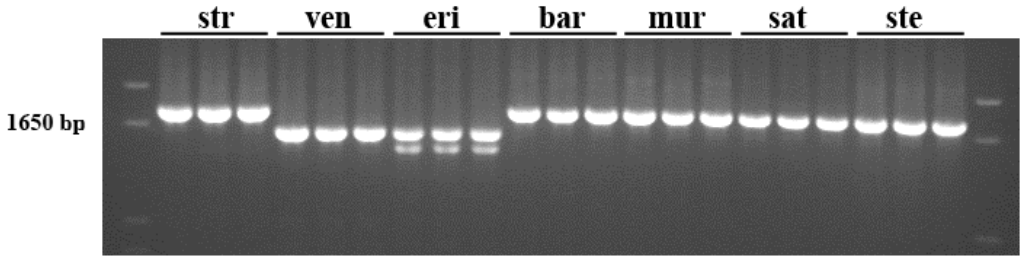


Figure 5. Amplification of the complete 5' external transcribed spacer obtained with primer combination ETS1_fow and 18S. Lane 1 and 23 – molecular marker 1Kb+; 2 to 4 – *A. strigosa*; 5 to 7 – *A. ventricosa*; 8 to 10 – *A. eriantha*; 11 to 13 – *A. barbata*; 14 to 16 – *A. murphyi*; 17 to 19 – *A. sativa* and 20 to 22 – *A. sterilis*.

All bands were cloned and sequenced using M13 and internal primers designed based on the partial sequences obtained. The full alignment can be seen in supplemental material (Appendix III) and Table 3 summarizes the highest and lowest sequence similarity between all clones sequenced. Amongst species with genome A, homology of 96 to 99% was obtained, even though *A. sativa* clone satII and *A. sterilis* clone steI sequences presented short duplications (in +650 to +681 and +595 to +621, respectively). The C-genome diploid species sequences were shown to be 99% similar. Comparison of the sequences of species with A-genome and C-genome diploid species revealed overall homology considerably lower. The lowest similarity of 70% was observed between *A. ventricosa* and the published sequence of *A. sativa* and the highest (76%) was detected between the shortest sequence of *A. eriantha* and *A. barbata*, *A. sativa* and *A. sterilis*.

Table 3: Similarity between complete external transcribed spacer sequences (ETS).

%	X74820	Str	Ven	EriL	EriS	Bar	Mur	Sat	Ste
X74820	100	97	70	71	75	97	98	96-97	96-99
Str		100	71	71	75	99	98	99	98
Ven			100	99	99	71-72	71	71-72	71-72
EriL				100	100	71	70-71	71-72	71
EriS					100	75-76	75	75-76	75-76
Bar						98	98	98-99	98-99
Mur							100	98	98
Sat								100	98-100
Ste									98

Note: EriL and EriS represent the sequences of the larger and shorter bands of *A. eriantha*, respectively. The similarity between the sequences of the C and A-genomes is highlighted in light grey.

Few SNPs were detected upstream the 18S which is a region prone to the formation of loops. Thus, the ability to form loops in the two genomes was evaluated through Mfold (Zuker, 2003) analysis revealing similar structures with similar free energies of -29.90 kcal/mol in species with A-genome sequences and of -30.50 kcal/mol in the C-genome diploid species. Therefore the SNPs do not affect loop conformation significantly.

Genome organization

To identify repetitive elements and further understand the organization of the 5' ETS, dotplots were constructed. All sequences of species with A-genome - *A. strigosa, A. barbata, A. murphyi, A. sativa* and *A. sterilis* - were fairly identical to the *A. sativa* sequence published. Thus, this sequence was compared with one sequence from *A. strigosa* as can be seen in Figure 6A where the dotplot marked all repeats reported previously: two repeats with ~150bp in the 300-600bp region corresponding to the C repeats; two short repeats in the 1100-1200

bp region, where the D repeats were described; and three repeats with approximately 120 bp in the 1200-1600bp region corresponding to the E repeats.

The organization of the C-genome diploid species, *A. ventricosa* and *A. eriantha*, is similar both in number and location of repeats for the ~1500bp sequence. The difference between these sequences and the 1400bp one present in *A. eriantha* was that the shorter variant lacks the first copy of the E repeat. Therefore, the *A. ventricosa* sequences were selected to represent the organization of the C-genome diploid species.

There are major differences in the organization of *A. ventricosa* when compared with the published *A. sativa* sequence as can be observed by the gaps only present in one of the axis of the dotplot (Figure 6B). Moreover, it is evident the presence of three E repeats in *A. ventricosa* sequence upstream the ones annotated in *A. sativa* sequence, as well as the existence of only one copy of the C repeat as observed in the partial sequences. The complete sequences also revealed the presence of only one E repeat closer to the 18S in *A. ventricosa,* where all A-genome species have three. The presence of two D repeats was also not represented in the dotplot (Figure 6C) which was confirmed in the alignment.

Comparison within the E repeats of *A. ventricosa* revealed that the three repeats closer to the C repeat are more similar (94-96%) while the E repeat closer to the 18S is 75 to 78% similar to the ones closer to the C repeat. The highest homology between the E repeats of *A. ventricosa* and *A. sativa* was 82% between repeat 2 and 3 of the published sequence and the repeat closest to the 18S in *A. ventricosa*. Surprisingly, there was higher similarity with two partial ETS sequences of other Aveninae sub-tribe species of *Helictotrichon* genus (91% - Accession numbers: KC899016.1 and GQ324269.1). The dotplot of *A. ventricosa* (Accession Number: KM586759) vs. GQ324269.1 sequence showed not only that the E repeat closer to the 18S is similar but revealed a higher conservation level as well (Figure 6D). Thus it seems that ETS of C-genome diploid species is closer to the *Helictotrichon* genus than to the remaining *Avena* species evaluated (*A. strigosa, A. barbata, A. murphyi, A. sativa* and *A. sterilis*). The other ETS sequences available from the Aveninae sub-tribe are from the genus *Cinna* and covers only 81% of the *A. ventricosa* E repeat sequence, instead of 100% of *Helictotrichon* and *Avena* sequences. However, the C-genome region covered is also more similar to *Cinna* (84%) than to *Avena sativa* (82%).

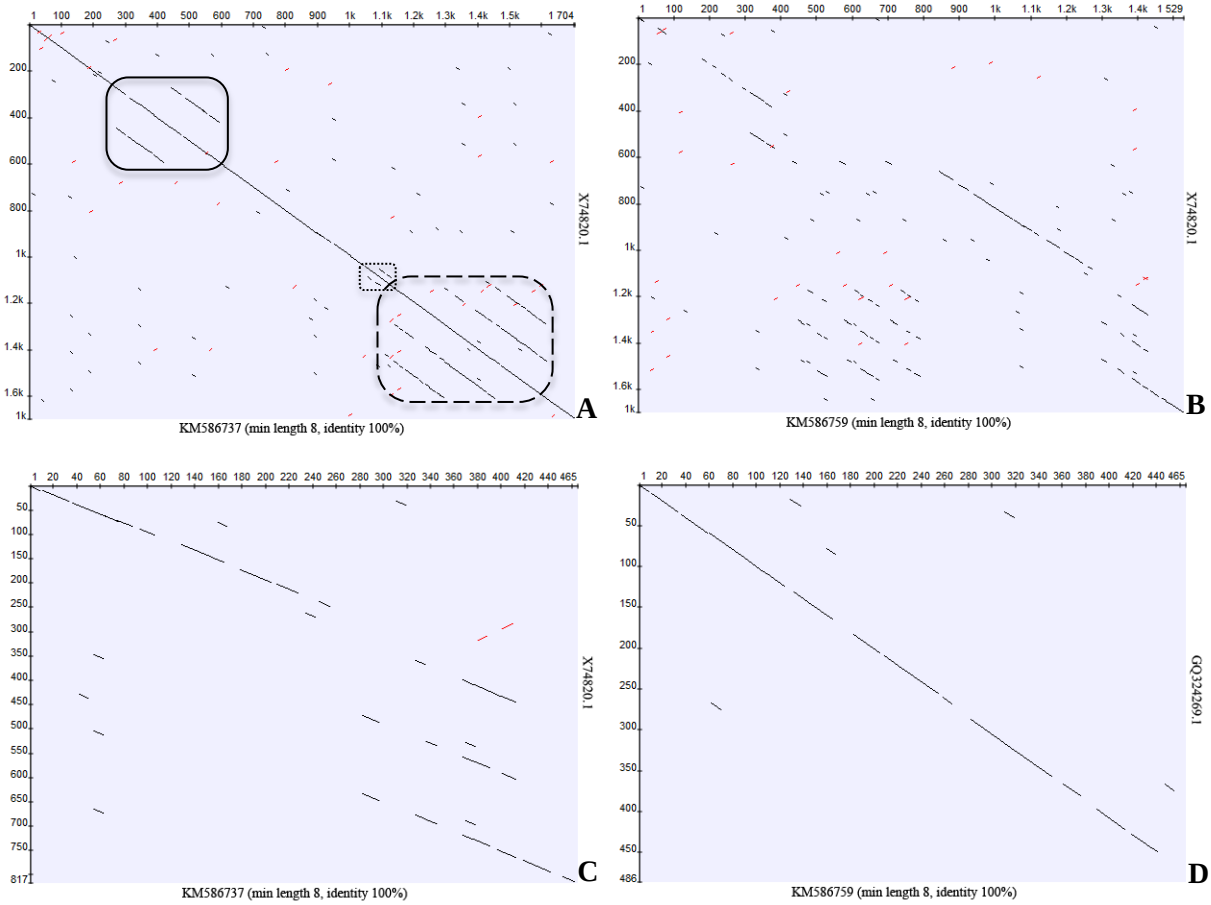


Figure 6. Dotplots of the 5' ETS. Comparison between the published sequence of *A. sativa* (Accession Number: X74820.1) and *A. strigosa* (Accession Number: KM586737 - A) or *A. ventricosa* (Accession Number: KM586759 - B). The +1011 to +1475 region of *A. ventricosa* was also compared with the same region of the published *A. sativa* (C) and with the homologous *H. sempervirens* sequence (Accession Number: GQ324269.1 - D). In A, solid outline highlights the C repeats, dot outlines D repeats and dash outline highlights the E repeats.

Evolution of the organizations A and C

In the attempt to verify the presence of sequences with organization C in the polyploid species with the C-genome – *A. murphyi*, *A. sativa* and *A. sterilis* –, the primers ETS4_fow and 18S were used to amplify a partial region of the 5' ETS in all species, since *in silico* analysis indicated that the ETS4_fow primer was absent in the A-genome. Amplification in all species in study obtained with those primers, as well as with primers ITS3 and ITS4 to amplify the internal transcribed spacer 2 (ITS2), can be seen in Figure 7. Both reactions were performed simultaneously and the control ITS2 reaction yielded a single band with ~380bp in all species, as expected (Figure 7, lanes 9 to 15).

The absence of amplification in the A-genome diploid *A. strigosa* with primers ETS4_fow and 18S confirmed that the ETS4_fow primer sequence is not present in this

species, confirming *in silico* data (Figure 7, lane 2). *A. ventricosa* and *A.eriantha* yielded the expected high intensity bands: a single band with ~1100bp in *A. ventricosa* and a band with similar size, as well as a ~1000bp band in *A. eriantha.* Interestingly, amplification fragments were detected in polyploid species with C-genome – *A. murphyi, A. sativa* and *A. sterilis* – and the dimensions of the bands observed were higher than those of the C-genome diploid species.

The sequences obtained from *A. ventricosa, A. murphyi, A. sativa* and *A. sterilis* revealed that: (i) in *A. ventricosa* the region amplified was the expected with three copies of the E repeat upstream the D repeat; (ii) the remaining species presented the organization C, with difference in sizes due to a higher number of E repeats upstream the D repeat. In fact, *A. sativa* and *A. sterilis* present four repeats and in *A. murphyi*, four, six and seven E repeats were observed (Alignment available in: https://www.dropbox.com/s/iov61m1s9gncyy2/Alignment%20ETS4_fow-18S.pdf?dl=0).

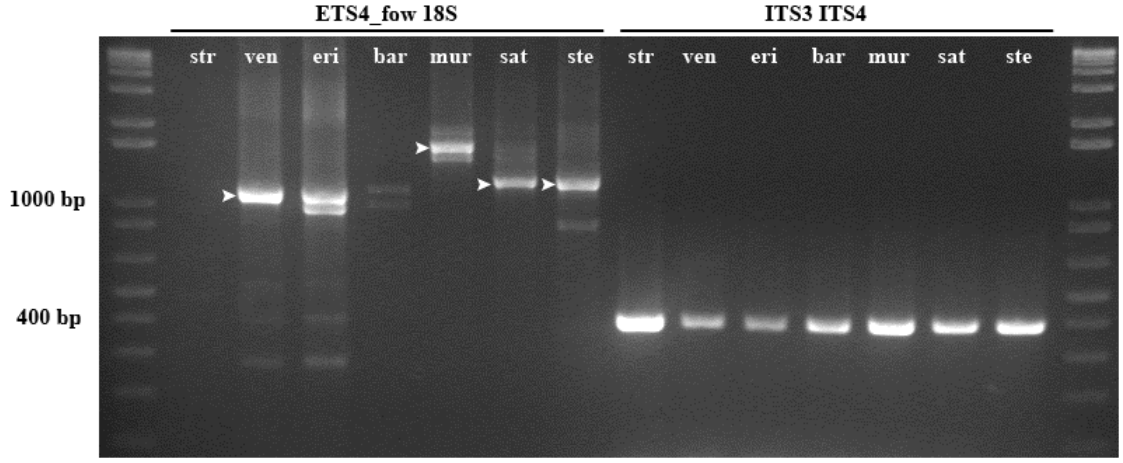


Figure 7. Simultaneous amplification of the ETS4_fow/18S region (A) and ITS2 (B). Lane 1 – molecular marker 1Kb+; 2 to 8 – primers ETS4_fow and 18S; 9 to 15 – primers ITS3 and ITS4; 16 – molecular marker 1Kb+. *A. strigosa* (2 and 9); *A. ventricosa* (3 and 10); *A. eriantha* (4 and 11); *A. barbata* (5 and 12); *A. murphyi* (6 and 13); *A. sativa* (7 and 14) and *A. sterilis* (8 and 15). Arrowheads mark the bands sequenced.

To attempt a deeper understanding of *Avena* species evolution based on the 5' ETS sequences this 45S rDNA domain was compared with other sequences available on NCBI from species of the same sub-tribe (*Helictotrichon sempervirens*, Accession Number: GQ324269.1); tribe (*Puccinellia bruggemannii* Accession Number: GQ283217.1) and the sub-family (*Triticum aestivum* Accession Number: X07841.1). This analysis revealed that the region from approximately +650 in *A. sativa* and from + 850 in *A. ventricosa* until the 18S is similar between the referred species. Table 4 summarizes the homology obtained with the

percentage of sequence covered in *H. sempervirens*, *P. bruggemannii* and *T. aestivum* being higher when compared with *A. ventricosa* than to the published *A. sativa* sequence.

Table 4. Sequence coverage and homology between *A. sativa* (Accession Number: X74820.1) and *A. ventricosa* (Accession Number: KM586759) with *T. aestivum* (Accession Number: X07841.1), *P. bruggemannii* (Accession Number: GQ283217.1) and *H. sempervirens* (Accession Number: GQ324269.1).

	A. sativa		***A. ventricosa***	
	Query coverage	**identities**	**Query coverage**	**identities**
T. aestivum	12%	79%	14%	79%
P. bruggemannii	77%	82%	85%	80%
H. sempervirens	95%	76%	97%	90%

The similarity levels obtained between *Avena* and *Triticum* or *Puccinellia* sequences are identical. However, 5' ETS sequences seem to be much more conserved between *Helictotrichon* and the C-genome diploid *A. ventricosa*, as suggested by the dotplot matrix (Figure 6D). Furthermore, the region from primer ETS1_fow to the TIS is also homologous between *Triticum* and *Avena* since both *Avena* sequences cover 97% of the wheat sequence and are 80 and 83% homologous to *A. sativa* and *A. ventricosa* respectively. The sequences of tribe Poeae and sub-tribe Aveninae available do not extend upstream of the transcription initiation site, therefore such analysis was not possible. Nonetheless, these results suggest that the organization C is more similar to other Poaceae species than the organization A.

Phylogenetic analysis

5'ETS sequences analysis

The phylogenetic trees constructed based on the 5' ETS complete sequences revealed that the C-genome diploid species *A. ventricosa* and *A. eriantha* form one group apart from the remaining species (Figure 8). The other tree branch includes one cluster with *A. sterilis* and *A. murphyi* and another one with *A. strigosa* and *A. barbata* being *A. sativa* isolated from the other two species.

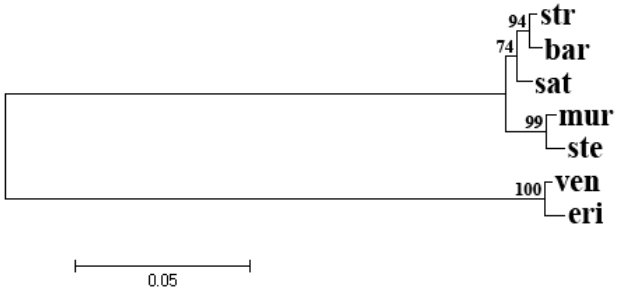


Figure 8. Neighbour-joining consensus tree based on the 5' ETS complete sequences.

Analysis through IRAP and REMAP

IRAP and REMAP methodologies were also performed to evaluate genomic diversity of all species analyzed and to obtain a phylogenetic tree based on genome-wide amplification. Two or three individuals of *A. ventricosa*, *A. eriantha*, *A. barbata* and *A. murphyi* were analyzed to assess intraspecific diversity. Since *A. strigosa*, *A. sativa* and *A. sterilis* had been previously evaluated (Tomás *et al.*, submitted) only one individual of each species was analyzed.

The lack of intravarietal variability was also demonstrated using IRAP and REMAP methodologies banding profiles through the analysis of distinct individuals of *A. ventricosa*, *A. eriantha*, *A. barbata* and *A. murphyi*, (Figure 9) as previously reported for *A. strigosa*, *A. sativa* and *A. sterilis*. Moreover, the patterns obtained in the other three species analyzed were similar to the ones reported previously (Tomás *et al.*, submitted). IRAP amplifies regions between the long terminal repeats (LTR) of retrotransposons and REMAP amplifies the same fragments, as well as between retrotransposons and microsatellites and between microsatellites, originating more complex patterns. IRAP banding profiles only allowed discrimination between *A. strigosa*, *A. barbata*, *A. murphyi*, *A. sativa* and *A. sterilis*. However, with the REMAP experiment it was possible to distinguish between all species, including the C-genome diploid species *A. ventricosa* and *A. eriantha.*

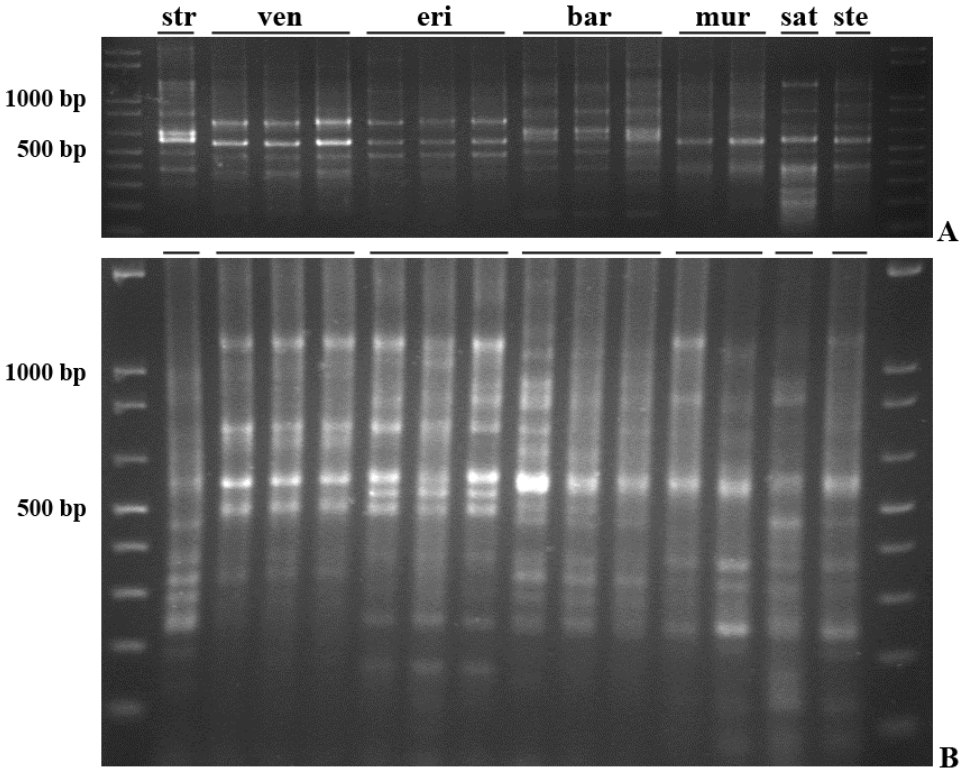


Figure 9. PCR banding profiles obtained with IRAP Nikita (A) and REMAP Nikita and (CA)9G (B). Lane 1 and 16 – molecular marker 1Kb+; 2 – *A. strigosa*; 3,4,5 – *A. ventricosa*; 6,7,8 – *A. eriantha*; 9,10,11 – *A. barbata*; 12,13 – *A. murphyi*; 14 – *A. sativa* and 15 – *A. sterilis*.

Binary matrixes of the banding profiles obtained with IRAP/REMAP were concatenated to generate a dendrogram. In Figure 10 it can be seen a branch constituted by two clusters: one including the A-genome diploid species *A. strigosa* and the tetraploid species *A. barbata* and the other encompassing the two hexaploid species, *A. sativa* and *A. sterilis*. Separated from these groups is another branch including the C-genome diploid species *A. ventricosa* and *A. eriantha* cluster and the allotetraploid *A. murphyi*. This division of species is in accordance with previous studies regarding *Avena* genus evolution that suggested that hexaploid species are closely related, as well as the species without the C-genome and indicate that AACC-genome *A. murphyi* retains more similarities with the C-genome diploid species than with the other group.

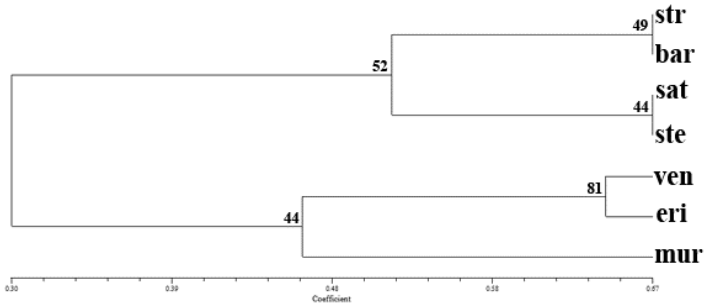


Figure 10. UPGMA clustering of the binary matrix from the IRAP and REMAP banding profiles.

IRAP/REMAP banding profiles vs 5'ETS sequences

The integrated analysis of both trees revealed similarities, specifically that *A. ventricosa* and *A. eriantha* are always in the same cluster and separated from all other species. *A. strigosa* and *A. barbata* are also clustered and *A. sativa* and *A. sterilis* are in the same cluster in the IRAP and REMAP tree (Figure 10) and, although separated in 5' ETS complete sequences tree (Figure 8), they are still fairly close. Concerning *A. murphyi*, the 5' ETS based analysis indicates that this allotetraploid species is closer to *A. sterilis*, clustered with all the other species with A-genome in a clade separated from the C-genome. However, the IRAP/REMAP tree reveals that *A. murphyi* is closer to the C-genome group although in a separated cluster.

Discussion

The main objective of the present work was to assess *Avena* genomic diversity, trying to enrich the knowledge concerning phylogenetic relations between distinct species with different ploidy levels and genomic constitutions. We focused mainly in the 45S rDNA unit since previous studies, in Poaceae species unraveled polymorphisms in the intergenic spacer (IGS) not only between varieties (Polanco and Pérez de la Vega, 1997) but also within populations (Flavell *et al.*, 1986) and individuals (Chou *et al.*, 1999). The amplification of the IGS of distinct individuals from all the *Avena* species analyzed revealed no intraspecific diversity, however, clear differences were identified between the banding patterns of most species analyzed. Polanco and Pérez de la Vega (1997) previously reported differences between *Avena* varieties, though more recently differentiation amongst species had not been possible through the amplification of the same region (Nikolodakis *et al.,* 2008). Although the banding profiles here obtained are less complex than the ones reported by Polanco and Pérez de la Vega (1997) they allow a higher discrimination in comparison with the study by Nikolodakis and colleagues (Nikolodakis *et al.,* 2008).

Shorter IGS fragments in comparison with previous reports were observed and sequencing one *A. sterilis* band revealed the presence of a truncated possibly nonfunctional IGS, since the transcription initiation site (TIS) is not present. Moreover, the analysis of the complete IGS of all species studied revealed no correlation between the level of ploidy and neither the length of the spacer nor the number of polymorphisms detected. However, repeat variations in the intergenic spacer are common (for recent examples see Gálian *et al.*, 2014; Inácio *et al.*, 2014) and have been proposed within accessions of *Avena sativa* (Polanco and Pérez de la Vega, 1997) thus supporting the differences observed in the IGS amplification. Polanco and Pérez de la Vega (1994) reported that C, D and E repeats are located in *Avena* 5' ETS and repeat number variation has been observed in Cucurbitaceae and Astereae (King *et al.*, 1993; Markos and Baldwin, 2002, respectively). Thus, our study was focused on this IGS domain to characterize repeat patterns variation within the 5' ETS still poorly studied in Poaceae.

The amplification of the partial 5' ETS unraveled an insertion upstream the transcription initiation site in most clones of *A. strigosa*, *A. sativa* and *A. sterilis* were the *Avena* TIS sequence TATAGTAGGG is located (Polanco and Pérez de la Vega, 1994). Equivalent regions were established as RNA polymerase I recognition sites in *Arabidopsis* (-55 to + 6, Doelling *et al.*, 1993 and Doelling and Pikaard, 1995), *Brassica* (-30 to +20, Sáez Vasquez and Pikaard, 2000) and *Triticum* (-113 to +15, Akhunov *et al.*, 2001). Thus, since *in*

vitro and transient expression essays suggested that mutations or deletions in TIS surrounding domains can block transcription or shift the region where it begins (Akhunov *et al.*, 2001), the differences observed in *A. strigosa, A. sativa* and *A. sterilis* may have functional significance.

Contrary to the results revealed by the amplification of the entire IGS, the amplification of the complete 5' ETS in all species with A-genome yielded only one fragment with the expected dimension (1750bp), suggesting that the differences observed in the length of total IGS are located in the non-transcribed spacer. Conversely, the 5' ETS is shorter in both C-genome diploid species *A. ventricosa* and *A. eriantha* and in *A. eriantha* the existence of polymorphisms in this region was verified by the amplification of two bands. On the other hand, neither hexa nor tetrapolyploids species with C-genome analyzed yielded the shorter fragments observed in the C-genome diploid species, which suggests their loss throughout evolution. Additionally, the detailed analysis of the 5' ETS sequences showed that all species with A-genome have the organization proposed by Polanco and Pérez de la Vega in *A. sativa* (1994). However, the C-genome diploid species revealed a different organization that, in comparison to the A-genome species, lacks one C and one D repeats, only possesses one E repeat downstream the D repeat and two or three of these repeats are located between the C and D sequences (Figure 11).

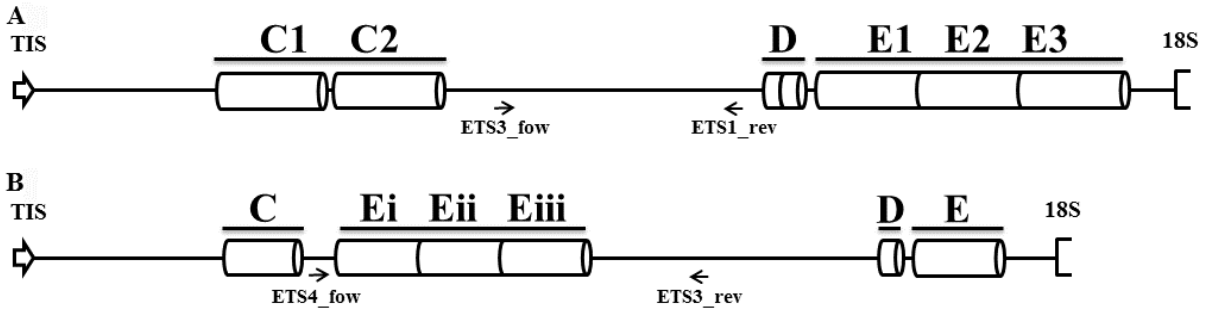


Figure 11. Representation of 45S rDNA 5' ETS repeat organization in species with A-genome (A, organization A) and in C-genome diploid species (B, organization C). Arrows represent the relative location of the internal primers used to sequence the complete 5' ETS.

There is no consensus regarding the *Avena* ancestral genome. C-genome diploid species have an asymmetrical karyotype, while the A-genome diploid species possesses a symmetrical karyotype, which is considered a primitive trait. However, the C-genome diploid species are morphologically very primitive (reviewed in Loskutov, 2008). The comparative analysis of our 5' ETS *Avena* sequences revealed that the organization of the C-genome diploid species is more similar to other Poaceae family *taxa* than to the A-genome organization suggesting that C-genome is closer to the ancestral Poaceae species. The detailed

analysis of *Avena* 5' ETS sequences seem to indicate additionally that the C repeat and surrounding sequence organization pattern is specific of *Avena*, since apart from the published sequence of *A. sativa*, no similar sequences were found through blast in NCBI database. However, the presence of the C repeats in other genera of the sub-tribe cannot be ruled out since no complete ETS sequences are available. Nonetheless, Polanco and Pérez de la Vega (1994) mentioned that the C repeats of *Avena* retain homology with the ones of wheat, hence it is not clear if there is a common origin between these repeats. Consequently, we suggest that the differences between the A and C-genomes of *Avena* arouse after the divergence of the *Avena* and *Helictotrichon* genera.

The 5' ETS repeats organization was further analyzed to understand the evolution of this region after genome divergence. The comparison of organization A ,which comprises all A-genome species – *A. strigosa, A. barbata, A. murphyi, A. sativa* and *A. sterilis* –, and organization C, that correspond to C-genome diploid species *A. ventricosa* and *A. eriantha*, revealed that the E repeats are more variable and the D repeats are the most similar. Also, considering the repeats homology within the organization A and C sequences, the succession of events involved in *Avena* species genome evolution may be proposed (Figure 12). The results presented suggest that one E repeat duplication in organization A and the duplication/translocation of the final E repeat to upstream the D repeat in organization C are the more ancient events. Afterwards, a duplication of the C sequence and a second duplication of the E repeat in organization A probably occurred, followed by the duplication of the D sequence. The two duplications of the translocated E repeat in genome C seem to have occurred more recently.

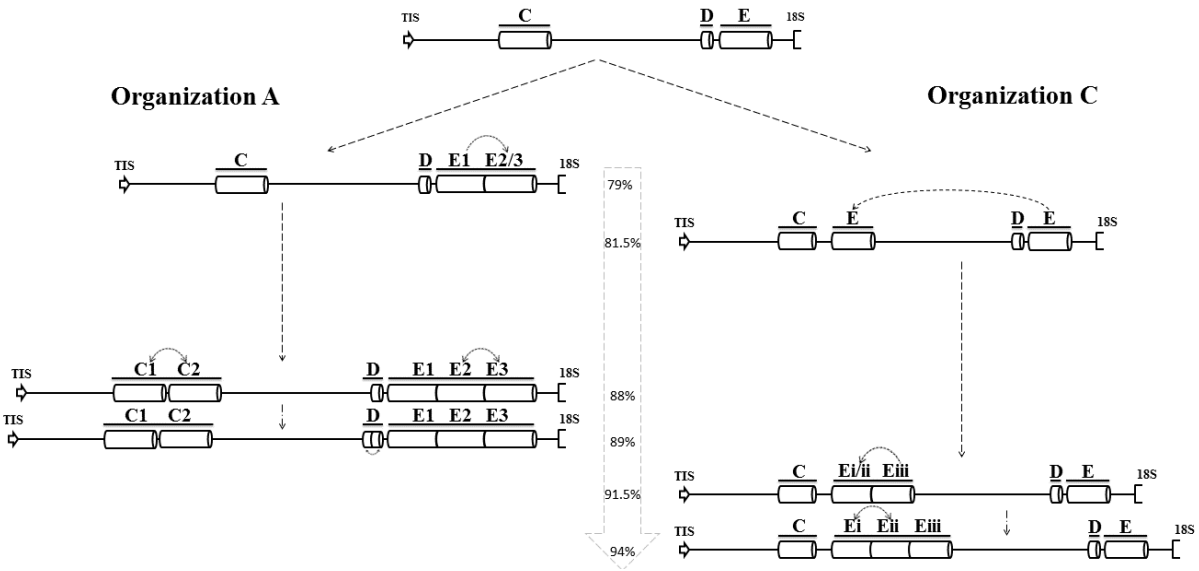


Figure 12. Representation of the proposed succession of events leading to the establishment of 45S rDNA 5' ETS organization A and C. The light grey arrow represents a relative time scale and the values within it, the similarity obtained by comparison of single nucleotide polymorphisms between repeats.

The 5' ETS could be a particularly useful option to enlighten the phylogenetic pathways of the genus *Avena* since high homology in the internal transcribed spaces in *Avena* has been extensively reported (Nikoloudakis *et al.*, 2008; Nikoloudakis and Katsiotis, 2008; Peng *et al.*, 2010; Rodionov *et al.*, 2005 and Tomás *et al.*, submitted) hindering an accurate evolutionary analysis (Tomás *et al.*, submitted). The phylogenetic tree based on the ETS sequences (Figure 8) unraveled a clear separation between the C-genome diploid species and the other *Avena* species analyzed in this work. However, the unexpected observation of two distinct organizations in this genus questioned the usefulness of the 5' ETS for phylogenetic inference in *Avena*. Therefore, an independent analysis was performed using IRAP and REMAP methodologies targeting retrotransposons and microsatellites flanking regions. Those molecular tools enable an accurate genome wide screening, previously described as a reliable tool to analyze intra and interspecific variation in *Avena* (Tomás *et al.*, submitted). The phylogenetic inference based on IRAP and REMAP genome-wide analysis is similar to the obtained with 5' ETS organizations. Only in *A. murphyi* the 5'ETS and IRAP/REMAP results are inconsistent since the rDNA sequences clustered this species with all other A-genome species analyzed while the former methodologies place *A. murphyi* closer to C-genome diploids species. Thus, although there are major differences in the 5' ETS organization patterns observed in *Avena*, these repetitive sequences proved to be suitable for interspecific phylogenetic inference.

Our results may also be relevant in the context of rDNA *loci* remodeling events that occur in polyploid species such as NOR loss (Maluszynska and Heslop-Harrison, 1993), gene silencing (Pikaard, 2000) and rDNA sequence homogenization (Nieto Feliner and Rosselló, 2012). Considering the number and location of NORs described for the species in this study, as summarized in Table 1, it is clear that in *Avena* species with C and A-genomes, C-genome origin NORs tend to be eliminated in polyploid species. On the other hand, the study of rDNA gene expression using silver-staining technic indicates that NOR silencing is not common in hexaploid species *A. sativa* and *A. byzantina* and in tetraploid species *A. murphyi* and *A. macrostachya* (Jellen *et al.*, 1994; Linares *et al.*, 1992; Fominaya *et al.*, 1988; Winterfeld and Röser, 2007 respectively). However, in AABB tetraploid species *A. barbata*, *A. vaviloviana* and *A. abyssinica*, inactive NORs have been reported (Badaeva *et al.*, 2010b; Irigoyen *et al.*,

2001; Fominaya *et al.*, 1988). Therefore it seems clear that in the genus *Avena*, polyploidy can induce a reduction of active NORs, either by NOR *loci* loss or by gene silencing.

Although there is a high copy number of rDNA in plant genomes, (Rogers and Bendich, 1987), these sequences are highly homogenized as expected by concerted evolution (reviewed in Nieto Feliner and Rosselló, 2012), a phenomenon observed in *Avena* in the present work. Though, the mechanisms implied in homogenization and restructuring events remain unknown, translocations, unequal crossing-over and transposition could be responsible. In the present work, even though some clones revealed duplications (satII Accession Number: KM586744 and steI Accession Number: KM586745) and eliminations (satIV Accession Number: KM586751), the amplification of the IGS and the 5'ETS analysis revealed a limited number of polymorphic bands. Though translocations have been reported in *A. sativa* (Irigoyen *et al.*, 2002), none of them are in the NOR harboring regions, thus, it seems that in *Avena* translocations are not related to the elimination of the C-genome NORs. On the other hand, no trace of a transposable element (TE) was identified on the C-genome sequences that could explain the duplication of the E repeat upstream of the D sequence, although interaction between TEs and rDNA has been suggested in *Arabidopsis* (Pontes *et al.*, 2004) and *Aegilops* (Raskina *et al.*, 2004 and 2008). Finally, unequal crossing-over may explain the two polymorphisms observed in the 5' ETS of *A. eriantha* as well as the tendency for E repeats accumulation.

Considering that differences in the A-genome are recent and widely spread and that the C-genome NORs tend to be eliminated in polyploids, it is tempting to speculate that the A-genome 5' ETS sequence may have a functional advantage. In fact, several keystones have been ascribed to that rDNA domain, namely the region recognized by the RNA polymerase I upstream the TIS (Doelling *et al.*, 1993, Doelling and Pikaard, Sáez Vasquez and Pikaard, 2000, Akhunov *et al.*, 2001), regions recognized by endonucleases in *Saccharomyces* (reviewed in Nazaar, 2004) also present in plants (Sáez-Vasquez *et al.*, 2004) and a loop just upstream the 18S (Polanco and Pérez de la Vega, 1994). Moreover, since after *Avena* genomes divergence both genomes accumulated E repeats, although in different regions, it seems that a high number of these sequences may be advantageous. However, the location of those sequences may be also important since the C-genome origin sequence is eliminated in *Avena* polyploid species.

An integrative view of the data obtained in this study indicates that *Avena* is certainly an interesting model to study ribosomal DNA evolution, to further comprehend rDNA processing in plants and to identify functional motifs that may be extremely useful to

understand mechanisms implied in restructuring events after polyploidization. The phylogenetic analysis with molecular markers IRAP and REMAP supports the use of 5' ETS that may also be crucial in the analysis of *Avena* evolution pathways to understand the genomic constitution of the ancestral of *A. barbata* as well as of most dissimilar members of the genus, like the geographically isolated *Avena insularis* and the C-genome tetraploid *Avena macrostachya.*

References

Akhunov, E., Chemeris, A., Kulikov, A., Vakhitov, V. (2001). Functional analysis of diploid wheat rRNA promoter by transient expression. Biochimica et Biophysica Acta 1522: 226-229.

Ananiev, E., Vales, M., Phillips, R., Rines, H. (2002). Isolation of A/D and C genome specific dispersed and clustered repetitive DNA sequences from *Avena sativa*. Genome 45: 431-441.

Badaeva, E., Shelukhina, O., Diederichsen, A., Loskutov, I., Pukhalskiy, V. (2010a) Comparative cytogenetic analysis of *Avena macrostachya* and diploid C-genome *Avena* species. Genome 53: 125-137.

Badaeva, E., Shelukhina, O., Goryunova, S., Loskutov, I., Pukhalskiy, V. (2010b) Phylogenetic Relationships of tetraploid AB-genome *Avena* species evaluated by means of cytogenetic (C-banding and FISH) and RAPD analyses. Journal of Botany 2010: 13 pages. Article ID 742307.

Badaeva, E., Shelukhina, O., Dedkova, A., Loskutov, I., Pukhalskiy, V. (2011). Comparative cytogenetic analysis of hexaploid *Avena* L. species. Russian Journal of Genetics 47 (6): 691-702.

Baumel, A., Ainouche, M., Kalendar, R., Schulman, A. (2002). Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* C.E. Hubbard (Poaceae). Mol Biol Evol 19 (8): 1218-1227.

Bento, M., Tomás, D., Viegas, W., Silva, M. (2013). Retrotransposons represent the most labile fraction for genomic rearrangements in polyploid plant species. Cytogenetic and Genome Research: 1-9. DOI: 10.1159/000353308.

Caperta, A.D., Neves, N., Morais-Cecílio, L., Malhó, R., Viegas, W. (2002) Genome restructuring in rye affects the expression, organization and disposition of homologous rDNA loci. Journal of Cell Science 115:2839-2846.

Chou, C-H., Chiang, Y-C., Chiang, T-Y. (1999). Within- and between-individual length heterogeneity of the rDNA-IGS in *Miscanthus sinensis* var. *glaber* (Poaceae): Phylogenetic analyses. Genome 42: 1088-1093.

Cluster, P., Allard, R. (1995). Evolution of ribosomal DNA (rDNA) genetic structure in colonial Californian populations of *Avena barbata.*Genetics 139:941-954.

Dice, L. (1945). Measures of the Amount of Ecologic Association between Species. Ecology 26: 297-302.

Doelling, J., Gaudino, R., Pikaard, C. (1993). Functional analysis of *Arabidopsis thaliana* rRNA gene and spacer promoters by transient expression. Proc. Natl. Acad. Sci. USA 90: 7528-7532.

Doelling, J., Pikaard, C. (1995). The minimal ribosomal RNA gene promoter of *Arabidopsis thaliana* includes a critical element at the transcription initiation site. The Plant Journal 8: 683-692.

Don, R., Cox, P., Wainwright, B., Baker, K., Mattick, J. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Research 19 (14): 4008.

Flavell, R., O'Dell, M., Sharp, P., Neto, E., Beiles, A. (1986). Variation in the intergenic spacer of ribosomal DNA of wild wheat, *Triticum dicoccoides*, in Israel. Mol. Biol. Evol 3 (6): 547-558.

Fominaya, A., Veja, C., Ferrer, E. (1988). C-banding and nucleolar activity of tetraploid *Avena* species. Genome 30: 633-638.

Fominaya, A., Hueros, G., Loarce, Y., Ferrer, E. (1995). Chromosomal distribution of a repeated DNA sequence from C-genome heterochromatin and the identification of a new ribosomal DNA locus in the *Avena* genus. Genome 38: 548-557;

Fu. Y., Chong, J., Fetch, T., Wang, M. (2007). Microsatellite variation in *Avena sterilis* oat germplasm. Theor Appl Genet 114: 1029-1038.

Fu, Y., Williams, D. (2008). AFLP variation in 25 *Avena* species. Theor Appl Genet 117:333-342.

Galián, J., Rosato, M., Rosselló, J. (2014). Incomplete sequence homogenization in 45S rDNA multigene families: intermixed IGS heterogeneity within the single NOR locus of the polyploid species *Medicago arborea* (Fabaceae). Annals of Botany 114 (2): 243-51.

Goffreda, J., Burnquist, W., Beer, S., Tansley, S., Sorrells, M. (1992). Application of molecular markers to assess genetic relationships among accessions of wild oat, *Avena sterilis*. Theor Appl Genet 85: 146-151.

Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41: 95-98.

He, X., Bjørnstad, Å. (2012). Diversity of North European oat analyzed by SSR, AFLP and DArT markers. Theor Appl Genet 125: 57-70.

Higo, K., Ugawa, Y., Iwamoto M., Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Research 27 (1): 297-300.

Inácio, V., Rocheta, M., Morais-Cecílio, L. (2014) Molecular organization of the 25S–18S rDNA IGS of *Fagus sylvatica* and *Quercus suber*: A comparative analysis. PLoS ONE 9(6): e98678.

Irigoyen, M., Loarce, Y., Linares, C., Ferrer, E., Leggett, M., Fominaya, A. (2001). Discrimination of the closely related A and B genomes in AABB tetraploid species of *Avena*. Theor Appl Genet 103: 1160-1166.

Irigoyen, M., Linares, C., Ferrer, E., Fominaya, A. (2002). Fluorescence *in situ* hybridization mapping of *Avena sativa* L. cv. SunII and its monosomic lines using cloned repetitive DNA sequences. Genome 45: 1230–1237.

Jellen, E., Phillips, R., Rines, H. (1994). Chromosomal localization and polymorphisms of ribosomal DNA in oat (*Avena* spp.). Genome 37:23-32.

Kalendar, R., Grob, T., Regina, M., Suoniemi, A., Schulman, A. (1999). IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98: 704-711.

Kalendar, R., Flavell, A., Ellis, T., Sjakste, T., Moisy, C., Schulman, A. (2011). Analysis of plant diversity with retrotransposon-based molecular markers. Heredity 106: 520-530.

Katsiotis, A., Hagidimitriou, M., Heslop-Harrison, J. (1997). The close relationship between the A and B genomes in *Avena* L. (Poaceae) determined by molecular cytogenetic analysis of total genomic, tandemly and dispersed repetitive DNA sequences. Annals of Botany 79: 103-109.

Katsiotis, A., Loukas, M., Heslop-Harrison, J. (2000). Repetitive DNA, Genome and Species Relationships in *Avena* and *Arrhenatherum* (Poaceae). Annals of Botany 86: 1135-1142.

King, K., Torres, R., Zentgraf, U., Hemleben, V. (1993). Molecular Evolution of the intergenic spacer in the nuclear ribosomal RNA genes of Cucurbitaceae. J Mol Evol 36: 144-152.

Koressaar, T., Remm, M. (2007). Enhancements and modifications of primer design program Primer3 Bioinformatics 23 (10): 1289-1291.

Larkin, MA., Blackshields, G., Brown, NP., Chenna, R., McGettigan, PA., McWilliam, H., Valentine, F., Wallace, IM., Wilm, A., Lopez, R., Thompson, JD., Gibson, TJ., Higgins, DC. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23 (21): 2947–2948.

Lawrence, P., Frey, K. (1975). Backcross variability for grain yield in oat species crosses (*Avena sativa* L. x *A. sterilis* L.). Euphytica 24: 77-85.

Leišová, L., Kučera, L., Dotlačil, L. (2007). Genetic resources of barley and oat characterised by microsatellites. Czech J. Genet. Plant Breed 43 (3): 97-104.

Li, C., Rossnagel, B., Scoles, G. (2000). The development of oat microsatellite markers and their use in identifying relationships among Avenaspecies and oat cultivars. Theor Appl Genet 101:1259-1268.

Linares, C., Vega, C., Ferrer, E., Fominaya, A. (1992). Identification of C-banded chromosomes in meiosis and the analysis of nucleolar activity in *Avena byzantina* C. Koch cv 'Kanota'. Theor Appl Genet 83: 650-654.

Linares, C., Gonzalez, J., Ferrer, E., Fominaya, A. (1996). The use of double fluorescence *in situ* hybridization to physically map the positions of 5s rDNA genes in relation to the chromosomal location of 18s-5.8s-26s rDNA and a C genome specific DNA sequence in the genus *Avena*. Genome 130(2): 535-541.

Linares, C., Ferrer, E., Fominaya, A. (1998). Discrimination of the closely related A and D genomes of the hexaploid oat *Avena sativa* L. Proc Natl Acad Sci USA 95: 12450-12455.

Linder, C., Goertzen, L., Heuvel, B., Francisco-Ortega, J., Jansen, R. (2000). The complete external transcribed spacer of 18S-26S rDNA: Amplification and phylogenetic utility at low taxonomic levels in Asteraceae and closely allied families. Molecular Phylogenetics and Evolution 14 (2): 285-303.

Loskutov, I. (2008). On evolutionary pathways of *Avena* species. Genet Resour Crop Evol 55: 211-220.

Loskutov, I. Rines, H. (2011). *Avena*. In Kole, C. (Ed.) Wild Crop Relatives: Genomic and Breeding Resources, Cereals. Springer Berlin Heidelberg, Pp. 109-183.

Maluszynska, J., Heslop-Harrison, J. (1993). Physical mapping of rDNA loci in *Brassica* species.Genome 36: 774–781.

Markos, S., Baldwin, B. (2002). Structure, molecular evolution, and phylogenetic utility of the 5' region of the external transcribed spacer of 18S-26S rDNA in *Lessingia* (Compositae, Astereae). Molecular Phylogenetics and Evolution 23: 214-228.

Morikawa, T., Nishihara, M. (2009). Genomic and polyploid evolution in genous *Avena* as revealed by RFLPs of repeated DNA sequences. Genes Genet. Syst. 84: 199-208.

Nazar, R. (2004). Ribosomal RNA processing and ribosome biogenesis in eukaryotes. Life 56(8): 457-465.

Nieto Feliner, G., Rosselló, J. (2012). Concerted Evolution of Multigene Families and Homoeologous Recombination . In Wendel, J., Greilhuber, J., Dolezel, J., Leitch, I. (eds.), Plant Genome Diversity Volume 1. Springer Vienna. Pp. 171-193.

Nikoloudakis, N., Katsiotis, A. (2008). The origin of the C-genome and cytoplasm of *Avena* polyploids. Theor Appl Genet 117: 273-281.

Nikoloudakis, N., Skaracis, G., Katsiotis, A. (2008). Evolutionary insights inferred by molecular analysis of the ITS1 – 5.8S – ITS2 and IGS *Avena* sp. Sequences. Molecular Phylogenetics and Evolution 46: 102-115.

Okonechnikov, K., Golosova, O., Fursov, M. (2012). Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28: 1166–1167.

Peng, Y., Wei, Y., Baum, B., Zheng, Y. (2008). Molecular diversity of the 5S rRNA gene and genomic relationships in the genus *Avena* (Poaceae: Aveneae). Genome 51:137-154.

Peng, Y., Baum, B., Ren, C., Jiang, Q., Chen, G., Zheng, Y., Wei, Y. (2010). The evolutionary pattern of rDNA ITS in *Avena* and phylogenetic relationship of the *Avena* species (Poaceae: Aveneae). Hereditas 147: 183-204.

Pikaard, C. (2000). The epigenetics of nucleolar dominance. Trends in Genetics 16 (11): 495-500.

Polanco, C., Pérez de la Vega, M. (1994). The structure of the rDNA intergenic spacer of *Avena sativa* L.: a comparative study. Plant Molecular Biology 25: 751-756.

Polanco, C., Pérez de la Vega, M. (1997). Intergenic ribosomal spacer variability in hexaploid oat cultivars and landraces. Heredity 78:115-123.

Pontes, O., Neves, N., Silva, M., Lewis, M., Madlung, A., Comai, L., Viegas, W., Pikaard, C. (2004). Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid Arabidopsis suecica genome. PNAS - Proceedings of the National Academy of Sciences of the United States of America 101: 18240-18245.

Raskina, O., Belyayev, A., Nevo, E. (2004). Quantum speciation in *Aegilops*: molecular cytogenetic evidence from rDNA clusters variability in natural populations. Proc Natl Acad Sci USA 101: 14818–14823.

Raskina, O., Barber, J., Nevo, E., Belyayev, A. (2008). Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. Cytogenet Genome Res 120:351–357.

Richard, G., Kerrest, A., Dujon, B. (2008). Comparartive genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiology and Molecular Biology Reviews 72 (4): 686-727.

Rodionov, A., Tyupa, N., Kim, E., Machs, E., Loskutov, I. (2005). Genomic configuration of the autotetraploid oat species *Avena macrostachya* inferred from comparative analysis of ITS1 and ITS2 sequences: on the oat karyotype evolution during the early events of the *Avena* species divergence. Russian Journal of Genetics 41 (5): 518-528.

Rogers, S., Bendich, A. (1987). Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. Plant Molecular Biology (9): 509-520.

Rohlf, F. (2000). NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1.

Sáez-Vásquez, J., Pikaard, C. (2000). RNA polymerase I holoenzyme-promoter interactions. Journal of Biological Chemistry 275: 37173-37180.

Sáez-Vásquez, J., Caparros-Ruiz, D., Barneche, F., Echeverría, M. (2004). A plant snoRNP complex containing snoRNAs, fibrillarin, and nucleolin-like proteins is competent for both rRNA gene binding and Pre-rRNA processing in vitro. Mol. Cell. Biol. 24 (16): 7284-7297.

Sáez-Vásquez, J., Echeverría, M. (2007). Polymerase I transcription. In Grasser, K. (Ed.) Annual plant reviews, Regulation of Transcription in Plants. Blackwell Publishing. Pp. 162-183.

Strychar, R. (2011). World Oat Production, Trade, and Usage. In Webster, F., Wood, P. (Eds) Oats: Chemistry and Technology, 2nd ed. 3340 Pilot Knob Road, St. Paul, Minnesota 55121, U.S.A. (AACC International, Inc.). Pp 1-10.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013).MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular biology and evolution 30: 2725-2729.

Thompson, J., Higgins, D., Gibson, T. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22 (22): 4673-4680.

Tomás, D., Rodrigues, J., Veloso, M., Viegas, W., Silva, M. Genomic diversity of *Avena* species from Portugal. Submitted to Plant Cell Reports.

Untergrasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B., Remm, M., Rozen, S. (2012). Primer3 - new capabilities and interfaces. Nucleic Acids Research 40 (15): e115.

White, T., Bruns, T., Lee, S., Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In Innis, M., Gelfand, D., Sninsky, J., White, T. (Eds.) PCR Protocols: A Guide to Methods and Applications. Academic Press, Inc. Pp. 315-322.

Winterfeld, G., Röser, M. (2007). Disposition of ribosomal DNAs in the chromosomes of perennial oats (Poaceae: Aveneae). Botanical Journal of the Linnean Society 155: 193-210.

Winterfeld, G., Döring, E., Röser, M. (2009). Chromosome evolution in wild oat grasses (Aveneae) revealed by molecular phylogeny. Genome 52: 361-380.

Yap, I., Nelson, R. (1996). WinBoot: A program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31 (13), 3406-3415.

Appendix

Appendix I. List of accession numbers of the sequences obtained in this study.

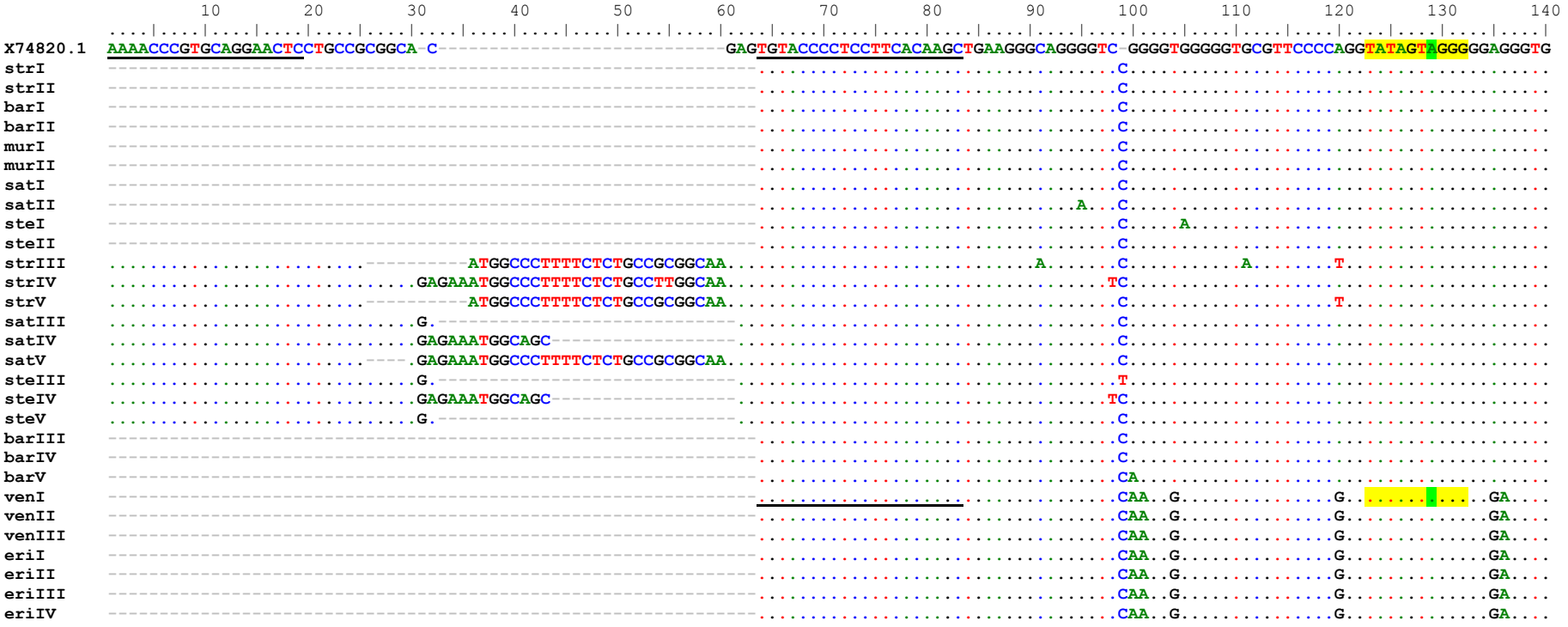
Clone Description	Accession Number
Avena strigosa, 45S rDNA external transcribed spacer, complete sequence, clone I	KM586737
Avena strigosa, 45S rDNA external transcribed spacer, complete sequence, clone II	KM586738
Avena strigosa, 45S rDNA external transcribed spacer, partial sequence, clone III	KM586747
Avena strigosa, 45S rDNA external transcribed spacer, partial sequence, clone IV	KM586748
Avena strigosa, 45S rDNA external transcribed spacer, partial sequence, clone V	KM586749
Avena ventricosa, 45S rDNA external transcribed spacer, complete sequence, clone I	KM586759
Avena ventricosa, 45S rDNA external transcribed spacer, complete sequence, clone II	KM586760
Avena ventricosa, 45S rDNA external transcribed spacer, partial sequence, clone III	KM586761
Avena ventricosa, 45S rDNA external transcribed spacer, partial sequence, clone IV	KM586772
Avena ventricosa, 45S rDNA external transcribed spacer, partial sequence, clone V	KM586773
Avena eriantha, 45S rDNA external transcribed spacer, complete sequence, clone I	KM586762
Avena eriantha, 45S rDNA external transcribed spacer, complete sequence, clone II	KM586763
Avena eriantha, 45S rDNA external transcribed spacer, complete sequence, clone III	KM586764
Avena eriantha, 45S rDNA external transcribed spacer, complete sequence, clone IV	KM586765
Avena barbata, 45S rDNA external transcribed spacer, complete sequence, clone I	KM586739
Avena barbata, 45S rDNA external transcribed spacer, complete sequence, clone II	KM586740
Avena barbata, 45S rDNA external transcribed spacer, partial sequence, clone III	KM586756
Avena barbata, 45S rDNA external transcribed spacer, partial sequence, clone IV	KM586757
Avena barbata, 45S rDNA external transcribed spacer, partial sequence, clone V	KM586758
Avena murphyi, 45S rDNA external transcribed spacer, complete sequence, clone I	KM586741
Avena murphyi, 45S rDNA external transcribed spacer, complete sequence, clone II	KM586742
Avena murphyi, 45S rDNA external transcribed spacer, partial sequence, clone III	KM586766
Avena murphyi, 45S rDNA external transcribed spacer, partial sequence, clone IV	KM586767
Avena murphyi, 45S rDNA external transcribed spacer, partial sequence, clone V	KM586768
Avena sativa, 45S rDNA external transcribed spacer, complete sequence, clone I	KM586743
Avena sativa, 45S rDNA external transcribed spacer, complete sequence, clone II	KM586744
Avena sativa, 45S rDNA external transcribed spacer, partial sequence, clone III	KM586750
Avena sativa, 45S rDNA external transcribed spacer, partial sequence, clone IV	KM586751
Avena sativa, 45S rDNA external transcribed spacer, partial sequence, clone V	KM586752
Avena sativa, 45S rDNA external transcribed spacer, partial sequence, clone VI	KM586769
Avena sativa, 45S rDNA external transcribed spacer, partial sequence, clone VII	KM586770
Avena sterilis, 45S rDNA external transcribed spacer, complete sequence, clone I	KM586745
Avena sterilis, 45S rDNA external transcribed spacer, complete sequence, clone II	KM586746
Avena sterilis, 45S rDNA external transcribed spacer, partial sequence, clone III	KM586753
Avena sterilis, 45S rDNA external transcribed spacer, partial sequence, clone IV	KM586754
Avena sterilis, 45S rDNA external transcribed spacer, partial sequence, clone V	KM586755
Avena sterilis, 45S rDNA external transcribed spacer, partial sequence, clone VI	KM586771
Avena sterilis, 45S rDNA intergenic spacer, clone VII	KM586774
Avena sterilis, 45S rDNA intergenic spacer, clone VIII	KM586775

Appendix II. Consensus sequence of the shorter intergenic spacer sequence of *A. sterilis.*

GACGACTTAAATACGCGACGGGGCATTGTAAGTGGCAGAGTGGCCTTGCTGCCACGATCCACTGAGATCCAGCCCCACGTCGCACGGATTCGTTACCTCGTGGTACTGGCACGTTT
TGTGCTCGGTGCTATCAAGGAAGCCTCGCTCTCGCTATTGGTTTCGGATGCCGCTCACGATAATGGTTAATGGCCCTTCTGGTTGCCGCTCACGATAAAGGTTAATAGCCCTTCTG
GTTGTCGCTAGTCCCACCTGAAAAGAAGCTTTCTGGATCTGAACGGCGGTAGTACGGTGTGTTGCATGTATTTCCCACAGTTTGTGGGGAGAAGCAACACGCTTGGAAACCTCTCT
AGTAGGAAAAGGAAAGCTTAGTCCCGTTCAAGTGCGACAACCGGACCGGCTATTGACCTCAAAACARGCTCTCTGGATCAGAACGGTCTTAGTACGGGGTGTTGCATGTATTCCCC
ACATTTTGTGGTAGGAGACGCATCACGCTGGAAATGTCTCTAGTAGGAGAAGGAAAGCTTAGTCCCGCTTAATAGCGACAACCGGAACGGCTGTTGACCCACCTCAAAAGAAGCCA
TCTGGATCAGAACGGACTTAGTACGGGGTGTTGCATGTATTCCCCACAGTTTGTGGTAGGAGACGCAACACGCTGGAAATGTCTCTAGTAGGTGAAGGAAAGCTTAGTCCCGATGA
ATTGCGACAACCGGACCGGCTGTTGTCTGCCTGCAAGGGCGGATGACTACCGTCGCTGGACGTCGAAGAGGACTCGCTACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCT

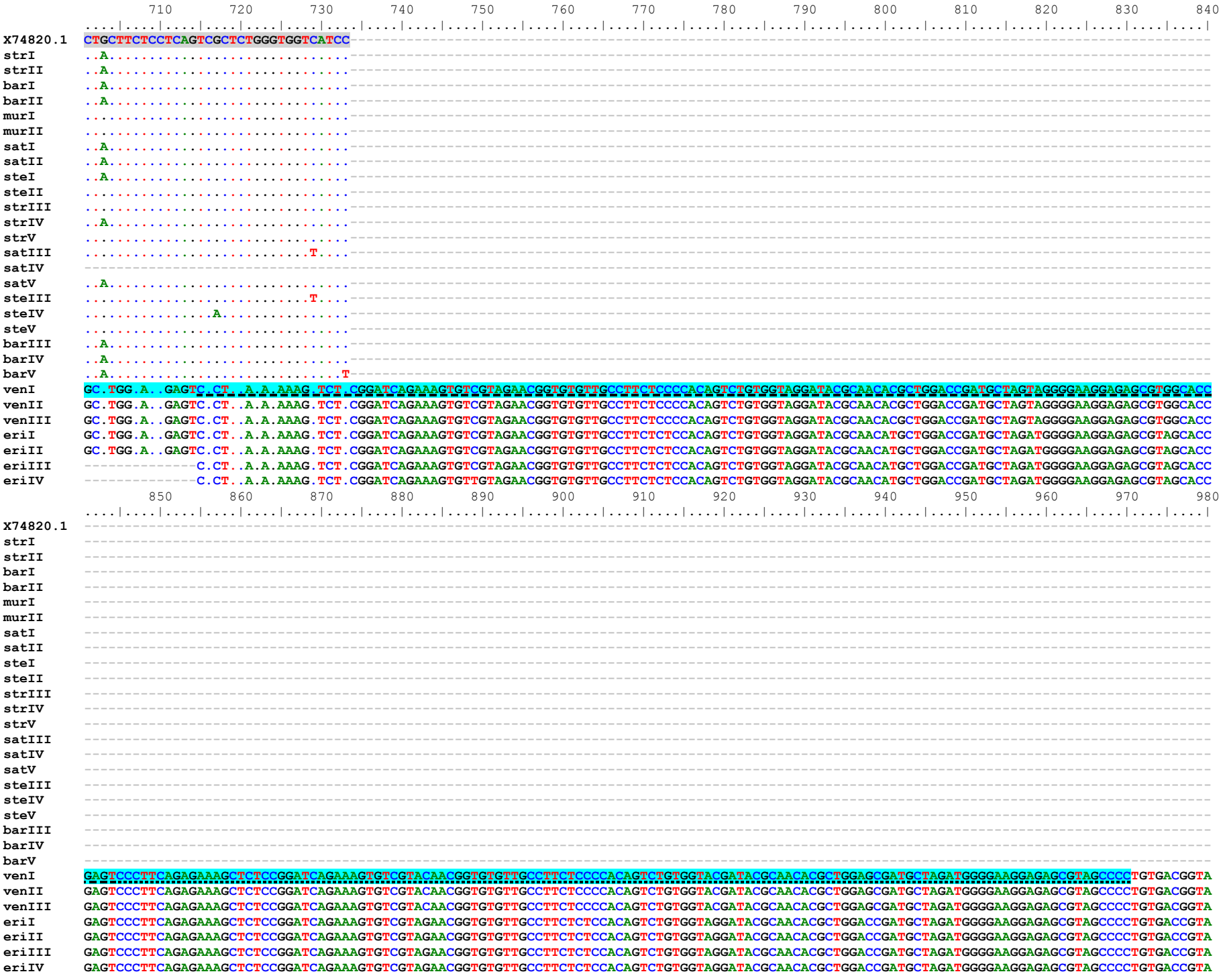
Note: The 25S sequence is underlined and the 18S sequence has the dash underline. Downstream of the 25S is homologous to the +975 to +1656 region of the *A. sativa* sequence (Accession Number: X74820.1). Highlighted in yellow is the nucleotide different between sequenced clones.

Appendix III. Alignment of partial and complete 5' ETS sequences with the published sequence of *A. sativa* between 2320 and 4098 bp (Accession Number: X74820.1).




150 160 170 180 190 200 210 220 230 240 250 260 270 280
X74820.1 GCCGAACCCCCC-------GGGGCGT-GCAAGGGTCGTGTTTTATGGGCAAAAACACGTGGATTTGAGTGGGA--TCCAAGAGGTTAGCGCAC-GCCCCTAGGCAAGACGTAGCGGTTGGCTCAGTCGGATGATGTTTC
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI ..GTGC.....TGGGTCTGT....A..T.GGT...A.....C.T.....A...TT....-------CC.A...A.A.T....C..GAG...G....C.---CA....A..---A.CG..
venII ..GTGC.....TGGGTCTGT....A..T.GGT...A.....C.T.....A...TT....-------CC.A...A.A.T....C..GAG...G....C.---CA....A..---A.CG..
venIII ..GTGC.....TGGGTCTGT....A..T.GGT...A.....C.T.....A...TT....-------CC.A...A.A.T....C..GAG...G....C.---CA....A..---A.CG..
eriI ..GTGC.....TGGGTCTGT....A..T.GGT...A.....C.T..G....A...TT....-------CC.A...A.A.T....C..GAG...G....C.---CA....A..---A.CG..
eriII ..GTGC.....TGGGTCTGT....A..T.GGT...A.....C.T..G....A...TT....-------CC.A...A.A.T....C..GAG...G....C.---CA....A..---A.CG..
eriIII ..GTGC.....TGGGTCTGT....A..T.GGT...A.....C.T..G....A...TT....-------CC.A...A.A.T....C..GAG...G....C.---CA....A..---A.CG..
eriIV ..GTGC.....TGGGTCTGT....A..T.GGT...A.....C.T..G....A...TT....-------CC.A...A.A.T....C..GAG...G....C.---CA....A..---A.CG..
290 300 310 320 330 340 350 360 370 380 390 400 410 420
X74820.1 CCTGTAACC--ATCACCGTAAA--CCGGTCCTTTA-GTATCCGTACGGGGATCTG--GGTCTGAGGTGTATTGCATGAGGTGTAGCACCATGAAATTGACGATTTGAGCTCCGCCGCGCAAGAACGTGTTAATCCGGTGT
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI .T..GCT..GT..T.T..A...AT.T....-..A.C.....A....A.CA..AT.TC.....----------.GC.T.GT.AC...A..T.....T....A..TT.....TC..T.CCTA
venII .T..GCC..GT..T.T..A...AT.T....-..A.C.....A....A.CA..AT.TC.....----------.GC.T.GT.AC...A..T.....T....A..TT.....TC..T.CCTA
venIII .T..GCC..GT..T.T..A...AT.T....-..A.C.....A....A.CA..AT.TC.....----------.GC.T.GT.AC...A..T.....T....A..TT.....TC..T.CCTA
eriI .T..GCC..GT..T.T..A...AT.T....-..A.C.....A....A.CA..AT.TC.....----------.GC.T.GT.AC...A..T.....T....A..TT.....TC..T.CCTA
eriII .T..GCC..GT..T.T..A...AT.T....-..A.C.....A....A.CA..AT.TC.....----------.GC.T.GT.AC...A..T.....T....A..TT.....TC..T.CCTA
eriIII .T..GCC..GT..T.T..A...AT.T....-..A.C.....A....A.CA..AT.TC.....----------.GC.T.GT.AC...A..T.....T....A..TT.....TC..T.CCTA
eriIV .T..GCC..GT..T.T..A...AT.T....-..A.C.....A....A.CA..AT.TC.....----------.GC.T.GT.AC...A..T.....T....A..TT.....TC..T.CCTA


430 440 450 460 470 480 490 500 510 520 530 540 550 560
X74820.1 CGTCCACGATACTTGGGAGCCGTTGCCCTGGTCATGTAAACGTTCCCAGTCGGTATGATTCCCCAAAAAGGTGATTGCCGCGTTACGTTTTTCGTGTTCCATGTGGGCTGCTTCTCCTCAGTCGCTCTGGGTGGTCATCC
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI
venII
venIII
eriI
eriII
eriIII
eriIV
570 580 590 600 610 620 630 640 650 660 670 680 690 700
X74820.1 ATGTTAACCACGGATAATGCGCTTGTCCGGTGTCGTCCACGATACTTGGGAGCCGCTGCCCTGGTCTCGTAAACGTTCCCAGTCGGTATGATTCCCCAAAAAGGTGATTGCCGCGTTACGTTTT CGTGTTCCATATGGG
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI GGTG.TTA..T.TTCG.GC..AG.CC.TTC.AAGAAAGCTCTCCGGATCAGAAA.TGT.GTA.AA.GGTGTGTT.CCTT.TC..CAC.GTC.GTGGT.GG.TACGC.ACA.GCT.GAC..A.GCTA.ATGGGG.AGGA.A
venII GGTG.TTA..T.TTCG.GC..AG.CC.TTC.AAGAAAGCTCTCCGGATCAGAAA.TGT.GTA.AA.GGTGTGTT.CCTT.TC..CAC.GTC.GTGGT.GG.TACGC.ACA.GCT.GAC..A.GCTA.ATGGGG.AGGA.A
venIII GGTG.TTA..T.TTCG.GC..AG.CC.TTC.AAGAAAGCTCTCCGGATCAGAAA.TGT.GTA.AA.GGTGTGTT.CCTT.TC..CAC.GTC.GTGGT.GG.TACGC.ACA.GCT.GAC..A.GCTA.ATGGGG.AGGA.A
eriI GGTG.TTA..T.TTCG.GC..AG.CC.TTC.AAGAAAGCTCTCCGGATCAGAAA.TGT.GTA.AA.GGTGTGTT.CCTT.TCC.CAC.GTC.GTGGT.GGGTACAC.ACA.GCT.GAC..A.GCTA..AGGGG.AGGA.A
eriII GGTG.TTA..T.TTCG.GC..AG.CC.TTC.AAGAAAGCTCTCCGGATCAGAAA.TGT.GTA.AA.GGTGTGTT.CCTT.TCC.CAC.GTC.GTGGT.GGGTACAC.ACA.GCT.GGC..A.GCTA..AGGGG.AGGA.A
eriIII GGTG.TTA..T.TTCG.GC..AG.
eriIV GGTG.TTA..T.TTCG.GC..AG.

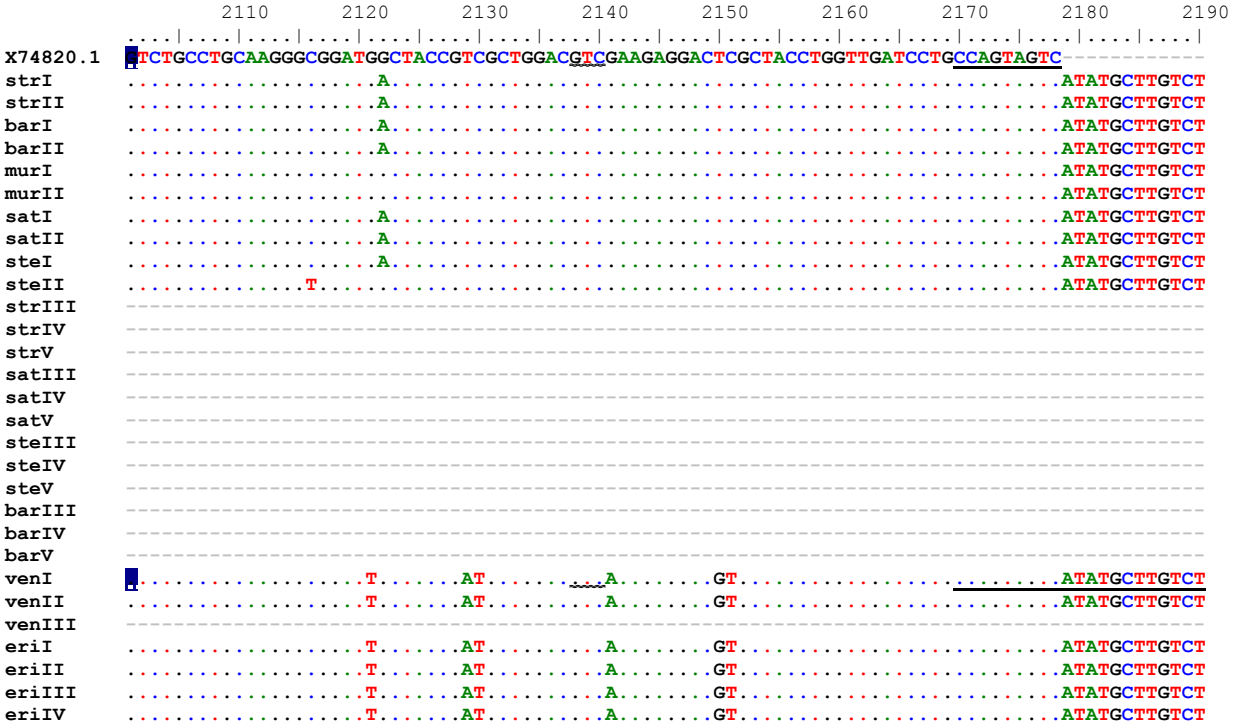

710 720 730 740 750 760 770 780 790 800 810 820 830 840
X74820.1 CTGCTTCTCCTCAGTCGCTCTGGGTGGTCATCC
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI GC.TGG.A..GAGTC.CT..A.A.AAAG.TCT.CGGATCAGAAAGTGTCGTAGAACGGTGTGTTGCCTTCTCCCCACAGTCTGTGGTAGGATACGCAACACGCTGGACCGATGCTAGTAGGGGAAGGAGAGCGTGGCACC
venII GC.TGG.A..GAGTC.CT..A.A.AAAG.TCT.CGGATCAGAAAGTGTCGTAGAACGGTGTGTTGCCTTCTCCCCACAGTCTGTGGTAGGATACGCAACACGCTGGACCGATGCTAGTAGGGGAAGGAGAGCGTGGCACC
venIII GC.TGG.A..GAGTC.CT..A.A.AAAG.TCT.CGGATCAGAAAGTGTCGTAGAACGGTGTGTTGCCTTCTCCCCACAGTCTGTGGTAGGATACGCAACACGCTGGACCGATGCTAGTAGGGGAAGGAGAGCGTGGCACC
eriI GC.TGG.A..GAGTC.CT..A.A.AAAG.TCT.CGGATCAGAAAGTGTCGTAGAACGGTGTGTTGCCTTCTCTCCACAGTCTGTGGTAGGATACGCAACATGCTGGACCGATGCTAGATGGGGAAGGAGAGCGTAGCACC
eriII GC.TGG.A..GAGTC.CT..A.A.AAAG.TCT.CGGATCAGAAAGTGTCGTAGAACGGTGTGTTGCCTTCTCTCCACAGTCTGTGGTAGGATACGCAACATGCTGGACCGATGCTAGATGGGGAAGGAGAGCGTAGCACC
eriIII ----------C.CT..A.A.AAAG.TCT.CGGATCAGAAAGTGTCGTAGAACGGTGTGTTGCCTTCTCTCCACAGTCTGTGGTAGGATACGCAACATGCTGGACCGATGCTAGATGGGGAAGGAGAGCGTAGCACC
eriIV ----------C.CT..A.A.AAAG.TCT.CGGATCAGAAAGTGTTGTAGAACGGTGTGTTGCCTTCTCTCCACAGTCTGTGGTAGGATACGCAACATGCTGGACCGATGCTAGATGGGGAAGGAGAGCGTAGCACC
850 860 870 880 890 900 910 920 930 940 950 960 970 980
X74820.1
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI GAGTCCCTTCAGAGAAAGCTCTCCGGATCAGAAAGTGTCGTACAACGGTGTGTTGCCTTCTCCCCACAGTCTGTGGTACGATACGCAACACGCTGGAGCGATGCTAGATGGGGAAGGAGAGCGTAGCCCCTGTGACGGTA
venII GAGTCCCTTCAGAGAAAGCTCTCCGGATCAGAAAGTGTCGTACAACGGTGTGTTGCCTTCTCCCCACAGTCTGTGGTACGATACGCAACACGCTGGAGCGATGCTAGATGGGGAAGGAGAGCGTAGCCCCTGTGACGGTA
venIII GAGTCCCTTCAGAGAAAGCTCTCCGGATCAGAAAGTGTCGTACAACGGTGTGTTGCCTTCTCCCCACAGTCTGTGGTACGATACGCAACACGCTGGAGCGATGCTAGATGGGGAAGGAGAGCGTAGCCCCTGTGACGGTA
eriI GAGTCCCTTCAGAGAAAGCTCTCCGGATCAGAAAGTGTCGTAGAACGGTGTGTTGCCTTCTCTCCACAGTCTGTGGTAGGATACGCAACACGCTGGACCGATGCTAGATGGGGAAGGAGAGCGTAGCCCCTGTGACCGTA
eriII GAGTCCCTTCAGAGAAAGCTCTCCGGATCAGAAAGTGTCGTAGAACGGTGTGTTGCCTTCTCTCCACAGTCTGTGGTAGGATACGCAACACGCTGGACCGATGCTAGATGGGGAAGGAGAGCGTAGCCCCTGTGACCGTA
eriIII GAGTCCCTTCAGAGAAAGCTCTCCGGATCAGAAAGTGTCGTAGAACGGTGTGTTGCCTTCTCTCCACAGTCTGTGGTAGGATACGCAACACGCTGGACCGATGCTAGATGGGGAAGGAGAGCGTAGCCCCTGTGACCGTA
eriIV GAGTCCCTTCAGAGAAAGCTCTCCGGATCAGAAAGTGTCGTAGAACGGTGTGTTGCCTTCTCTCCACAGTCTGTGGTAGGATACGCAACACGCTGGACCGATGCTAGATGGGGAAGGAGAGCGTAGCCCCTGTGACCGTA


990 1000 1010 1020 1030 1040 1050 1060 1070 1080 1090 1100 1110 1120
X74820.1 GCGGTAGGCACATTCCGTCGCCTCTAGTCCCTTCAGA GAAATGTTGCCCCAAGACGGTATCGTTATTGGTGTAACACCCGCGCGCCGTGGACACTCA
strI
strII
barI
barII
murI
murII
satI
satII TTGGT
steI CATTCCATCGCCTCTAGTCCCTTCAGA
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI ATTGTTGCCGA
venII ATTGTTGCCGA
venIII ATTGTTGCCGA
eriI ATTGTTGCCGA
eriII ATTGTTGCCGA
eriIII ATTGTTGCCGA
eriIV ATTGTTGCCGA
1130 1140 1150 1160 1170 1180 1190 1200 1210 1220 1230 1240 1250 1260
X74820.1 GGACGCCTTCGTGCAGGCCATTCCCCCGTGCTATGCGTGGGGGGAGGCTTGGTTGGCTTGACCGATGTGGATACGCGAGCGGATGAGTAGCTTTGGACCCGTGTCTGTTGTTA
strI
strII
barI
barII
murI
murII
satI
satII GTAACACCCGCGCGCCGTGGACACTCA
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI
venII
venIII
eriI
eriII
eriIII
eriIV


1270 1280 1290 1300 1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
X74820.1 GATCCCCCGTCCTCGTGCGGCCGACTAGCGACGCCGTGTCC GTCATTCGTGTGGCTTATGTTACTGTTGCCTTTCAAGTGCTTGCGTGCATGTACC GACCTACGG AAGTGGTGCTTT TACACGTTTGCCTCGCG C
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI
venII
venIII
eriI
eriII
eriIII
eriIV
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500 1510 1520 1530 1540
X74820.1 GG ACCTCTCGGTGTTCCGCTGTGGCCTAATGGCGCTTGCGGCGTTACCTCGTGGTACTGCGACGTTTCGTGCTCGGTGCTATCAAGGAAGCCTCGCTCTTGCTGTTGGTTTCGGATGCCGCTCACGATAATGGTTAAT
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI
venII
venIII
eriI
eriII
eriIII
eriIV


1550 1560 1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680
X74820.1
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI
venII
venIII
eriI
eriII
eriIII
eriIV
1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800 1810 1820
X74820.1
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI
venII
venIII
eriI
eriII
eriIII
eriIV


1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960
X74820.1
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI
venII
venIII
eriI
eriII
eriIII
eriIV
1970 1980 1990 2000 2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
X74820.1
strI
strII
barI
barII
murI
murII
satI
satII
steI
steII
strIII
strIV
strV
satIII
satIV
satV
steIII
steIV
steV
barIII
barIV
barV
venI
venII
venIII
eriI
eriII
eriIII
eriIV



Note: Primer sequence of ETS1_fow (from position 64 to 83), ETS2_fow (from position 1 to 19) and 18S (from position 2170 to 2190) are underlined.

The TIS is highlighted in yellow. TIS position +1 is highlighted in green.

C repeats are highlighted in grey. D repeats are highlighted in dark yellow: D1 is not underlined; D2 has a dash underline. E repeats are highlighted in blue: Ei is highlighted in light blue; Eii is highlighted in light blue and has a dash underline; Eiii is highlighted in light blue and has a dot underline. E1 is highlighted in dark blue; E2 is highlighted in dark blue and has a dash underline; E3 is highlighted in dark blue and has a dot underline.

Conserved sequences described previously are highlighted in pink (Polanco and Pérez de la Vega, 1994). Wave underlines mark duplications near the conserved sequences.

Double wave underlines the beginning and end of the loop formation sequence.